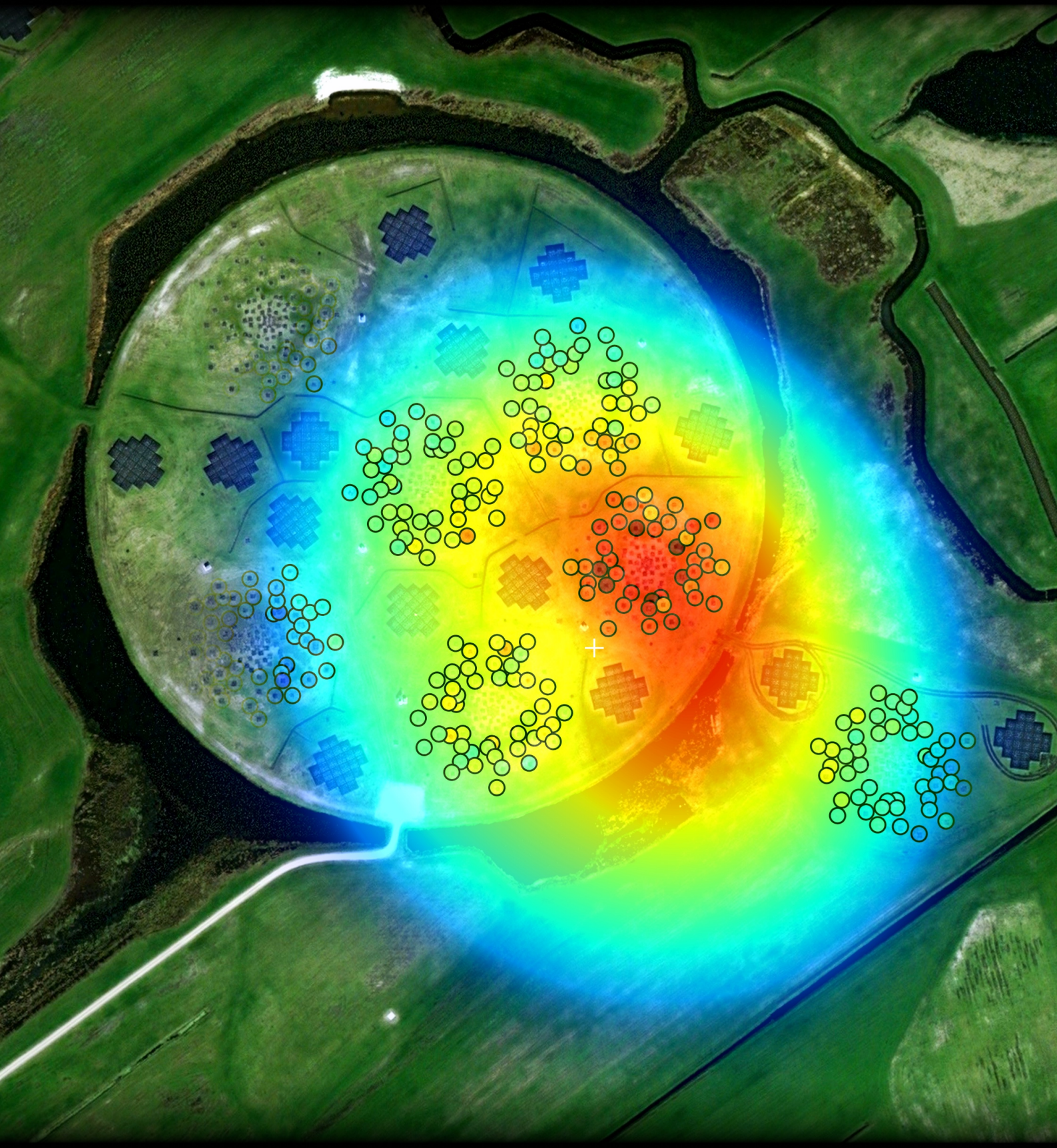
Cosmic rays at LOFAR

Precision measurements with a radio telescope



Arthur Corstanje

Cosmic rays at LOFAR

Precision measurements with a radio telescope

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,
volgens besluit van het college van decanen
in het openbaar te verdedigen op
maandag 23 september 2019
om 12:30 uur precies

door

Arthur Charles Willem Corstanje

geboren op 10 mei 1980
te Delft

Promotoren:	prof. dr. H.D.E. Falcke	
	prof. dr. J.R. Hörandel	
Manuscriptcommissie:	prof. dr. S.J. de Jong	
	prof. dr. A. Achterberg	
	prof. dr. R. Engel	Karlsruher Institut für Technologie (Duitsland)
	dr. M. Haverkorn	
	dr. J. Vink	Universiteit van Amsterdam

Cosmic rays at LOFAR
Precision measurements with a radio telescope
Thesis, Radboud University Nijmegen
Illustrated; with bibliographic information and Dutch summary
Cover: satellite image of the LOFAR core region (source: Microsoft Bing Maps),
with a simulated cosmic ray radio footprint
and corresponding LOFAR measurements
Cover layout by Ilse Modder
Printed by Gildeprint

ISBN: 978-94-6323-766-6

Science is a way of life. Science is a perspective. Science is the process that takes us from confusion to understanding in a manner that's precise, predictive and reliable - a transformation, for those lucky enough to experience it, that is empowering and emotional.

Brian Greene

Contents

Chapter 1

Introduction

In modern times, the study of the objects in the far universe is not limited to the use of light and other electromagnetic radiation to make images and spectra. Some of these objects, especially those with extreme or violent conditions such as e.g. supernova explosions and active galactic nuclei, are well known for their emission of high-energy particles, such as atomic nuclei, neutrinos, and high-energy gamma rays. Traveling at (nearly) the speed of light, and carrying energies up to more than a million times what is achievable in today's man-made particle accelerators, some arrive at Earth to be measured. The field of research which concerns these measurements is called *astroparticle physics*, as it is 'in between' astronomy and particle physics. The term 'cosmic rays' refers to atomic nuclei which arrive at the Earth's atmosphere.

This thesis contributes to the development of one particular method to measure high-energy cosmic rays, namely by detecting short radio-wavelength pulses on the ground.

1.1 Cosmic rays

Cosmic rays were discovered in 1912, when Victor Hess (Hess, 1912) made balloon measurements in which he found an increase of ionizing radiation with altitude, above about 1 km. At the time, the only known ionizing radiation came from radioactive processes at ground level, i.e. in the Earth's crust. Hence, he concluded that there was a newly discovered form of radiation, entering the atmosphere from outer space.

Later balloon experiments showed that these cosmic rays mainly consist of charged particles, being atomic nuclei ranging from protons to the heavy elements (Stanev, 2010). The primary particles enter the atmosphere at energies spanning a continuous spectrum from about 10^9 to 10^{20} eV. When a cosmic particle interacts with a nucleus in an air molecule, it creates secondary particles, which in turn also interact, producing a cascade of particles known as an *(extensive) air shower*. Measuring the energy and particle type of cosmic rays is possible through direct measurements in balloon or satellite experiments for energies up to about 10^{14} eV. Due to the steep fall-off of the spectrum, which reaches 1 particle per m^2 per year slightly above 10^{15} eV, and only 1 particle per km^2 per century at 10^{19} eV, measurements at higher energies are only feasible at large ground-based observatories. Here, the mass and energy of the incoming particle are inferred from the 'footprint' of secondary particles from the air shower, that arrive at the ground as a shower front from which also the incoming direction is inferred. Alternatively, methods such as fluorescence detection are used which image the shower evolution.

1.2 Production of air showers

At the first interaction of a cosmic nucleus with a nucleus in an air molecule, about half of its energy is converted to produce mesons, mainly pions. The target nucleus carries away the rest of the energy, allowing it to interact again further in the shower development.

The air shower can be separated into a hadronic, an electromagnetic, and a muonic part. The electromagnetic part arises from neutral pions (π^0) which decay

very quickly ($\tau \sim 10^{-16}$ s in co-moving frame) to two photons:

$$\pi^0 \to 2\,\gamma. \tag{1.1}$$

The photons are energetic enough to produce electron-positron pairs,

$$\gamma \to e^+ + e^-, \tag{1.2}$$

which again produce photons through bremsstrahlung, Compton scattering and annihilation of the positrons with electrons in the atmosphere. As long as the photons and leptons carry enough energy, above a critical energy of about 84 MeV (Matthews, 2005), this process continues. It is this part of the shower that produces the radio emission.

The hadronic shower develops through the charged pions π^+ and π^- which make up two-thirds of the produced pions, and have a much longer decay time, $\tau \sim 10^{-8}$ s. They decay into muons and neutrinos as

$$\pi^+ \to \mu^+ + \nu_\mu \tag{1.3}$$

$$\pi^- \to \mu^- + \bar{\nu_\mu}. \tag{1.4}$$

However, due to the long decay time they will often interact again before decaying, as long as the energy per particle, and the corresponding relativistic time dilation, are large enough.

As the shower develops, it reaches a point where the number of particles is maximal. The column density (g/cm^2) of traversed matter where this happens is referred to as $X_{\max}$, the depth of the shower maximum. The amount of mass in the atmosphere is such that this maximum is in fact reached, even for most cosmic rays up to the highest energies, at altitudes of roughly 2 to 6 km at moderate zenith angles.

The point of first interaction depends on the primary particle mass; e.g. an iron nucleus has a larger interaction cross section than a proton. Therefore, also the $X_{\max}$ point is reached higher up in the atmosphere for iron nuclei than for protons, on average. Following a simplified model by Heitler (1954) and Matthews (2005), it is seen that on average, $X_{\max}$ depends on atomic mass number A at a fixed primary energy E as

$$X^A_{\max} = X^p_{\max} - X_0 \ln(A), \tag{1.5}$$

where X_0 is the elongation rate for a proton, and $X^p_{\max}$ is the average $X_{\max}$ for protons. It is hereby assumed that a shower from a nucleus with mass number A can be described as a superposition of A proton showers, each carrying an energy E/A.

Measuring primary energy and $X_{\max}$ of air showers thus allows to infer their composition, at least statistically from a large sample.

1.3 Sources and composition of cosmic rays

One of the main open questions on cosmic rays is their origin: what produced them? Especially towards the high end of their energy spectrum, it is clear that acceleration to such energies requires extreme or violent conditions.

The most important acceleration mechanism is *Fermi acceleration* (Fermi, 1949; Bell, 1978), where the particles interact with plasma shock regions carrying strong and turbulent magnetic fields. Such plasma shocks are found e.g. in supernova remnants, active galactic nuclei, and in the lobes of radio galaxies. The acceleration does not happen in one interaction, but in a series of interactions, where at each interaction there is an average energy gain of

$$\frac{\Delta E}{E} \approx \frac{4}{3}\beta, \tag{1.6}$$

with $\beta = v/c$ the velocity of the shock.

This can only happen as long as the particle is magnetically contained in the source region. Thus, the limit for acceleration is given by the *Hillas criterion* of magnetic confinement (Hillas, 1984), giving the maximum attainable energy as

$$\frac{E_{\max}}{10^{15}\,\mathrm{eV}} \approx Z\,\frac{\beta}{2}\left(\frac{B}{1\,\mu\mathrm{G}}\right)\left(\frac{R}{1\,\mathrm{pc}}\right), \tag{1.7}$$

with the electric charge Z in units of the elementary charge, and β the plasma shock velocity. In the limit of relativistic shocks, with $\beta \to 1$, which is often considered for sources of particles with the highest energies, it follows from setting the particle's Larmor radius equal to the size R of the source region. The Hillas criterion is depicted in Fig. 1.1, for $\beta = 1$, showing the relation between required size and magnetic field for a $10^{20}\,\mathrm{eV}$ proton and iron nucleus. It shows that at these high

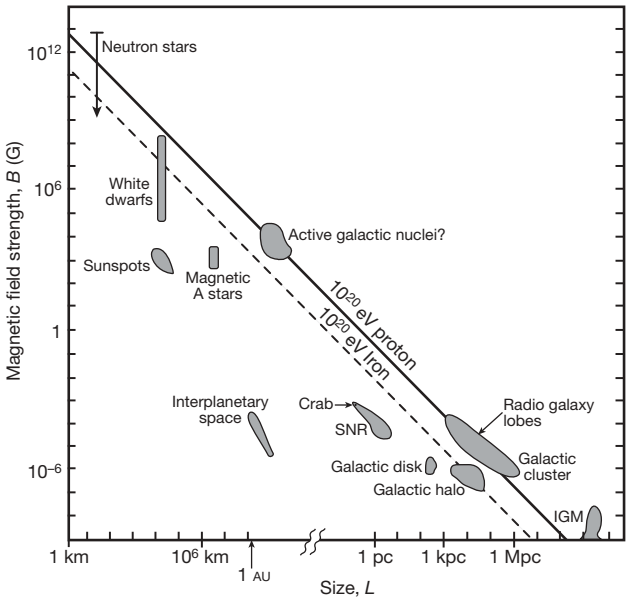


Figure 1.1: The Hillas plot, illustrating the required magnetic field strength as a function of the size of potential cosmic ray sources. Protons and iron nuclei can only be accelerated to an energy of 10^{20} eV in objects at or above the respective lines shown in this plot. Figure from Bauleo and Martino (2009).

energies and especially for protons, only large extragalactic sources, or perhaps active galactic nuclei are feasible candidates for the origin of these particles.

Importantly, for a given source the maximum energy is higher for heavier nuclei than for protons, up to a factor 26 for iron. So, when observing cosmic rays from a given type of source, we expect their composition to become heavier towards higher energies.

Due to the general presence of magnetic fields in the Galactic and intergalactic medium, one cannot trace back the cosmic rays to their sources, from their arrival directions. Only for protons or light nuclei (i.e. small charge Z) above 10^{19} eV, such a search is feasible. In this regime an anisotropy in the arrival directions has been found (Aab et al., 2017). It should be noted that also at lower energies, in the range

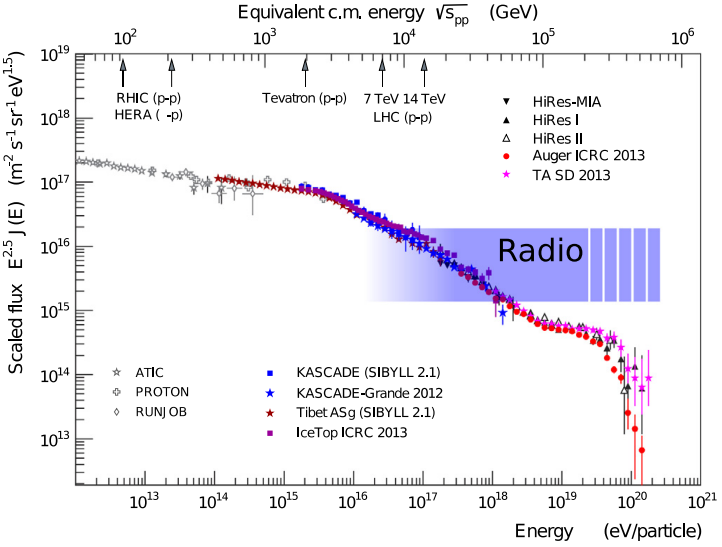


Figure 1.2: The cosmic ray energy spectrum, as measured by several experiments. Figure from Huege (2016), adapted from Engel et al. (2011).

2 to 70 TeV, small deviations from isotropy in the arrival directions have been found, e.g. in the High Altitude Water Cherenkov observatory (HAWC) (Abeysekara et al., 2018) and for the TeV to PeV regime in the IceCube observatory (Aartsen et al., 2016). The less abundant high energy gamma rays and neutrinos can be directly traced back, as they do not suffer deflection by magnetic fields.

Supernova remnants, which are present in our Galaxy, are natural candidates for particle acceleration up to about 10^{15} eV, while above 10^{18} eV the cosmic rays are believed to originate from extragalactic accelerators. Hence, in the spectrum there is a transition region, crossing from galactic to extragalactic origin. In the light of the Hillas criterion and its charge dependence, studying the composition of cosmic rays as function of energy (which is interesting in its own right) also gives information about their sources. The energy spectrum, although rather smooth, shows a number of features at high energies as seen in Fig. 1.2. The two most important ones are called the 'knee', around $3 \cdot 10^{15}$ eV where the spectral index drops from about -2.7 to -3.1, and the 'ankle' at about $4 \cdot 10^{18}$ eV where the spectral index rises

again (Blümer et al., 2009; Hörandel, 2003). At the highest energies, above about $5 \cdot 10^{19}$ eV, a steep cutoff is expected, as the primary particles may then produce pions in interactions with photons from the cosmic microwave background. In the rest frame of the primary particle, these photons reach the energy threshold of pion production. This is called the GZK-cutoff, after Greisen, Zatsepin and Kuzmin (Greisen, 1966; Zatsepin and Kuzmin, 1966).

Particles above the GZK-limit have been detected, especially by large-area experiments like the Pierre Auger Observatory. If they are protons, they are at the upper edge of the feasible sources in the Hillas diagram (Fig. 1.1). Also, through energy loss of the photo-pion production process, the distance of their origin must be constrained to within 10 to 100 Mpc. As from the Hillas criterion it is unclear if astrophysical accelerators can produce significant numbers of particles above 10^{20} eV, it cannot be straightforwardly concluded that the observed drop in the spectrum is only due to the GZK-cutoff.

Studying the mass composition is helpful here; when a natural limit to particle acceleration is more important, one expects to see a trend towards heavier composition near the cutoff. This would not be the case when the cutoff is purely due to the GZK effect.

1.4 Techniques for air shower detection and composition measurements

The most direct way to detect air showers is by using an array of particle detectors, ideally spread over a large area, that detects particles in coincidence when an air shower arrives. From particle detectors alone, the composition can be measured using the electron-to-muon ratio at the ground, for detectors capable of separating the electron and muon contributions. This is done e.g. in KASCADE-Grande (Apel et al., 2010), Yakutsk (Afanasiev et al., 2003), IceTop (Abbasi et al., 2013) and is planned for the Pierre Auger Observatory (Cataldi, 2017). By the superposition principle discussed in Section 1.2, higher-mass particles produce showers with relatively more muons, as they can be taken as A equivalent proton showers for primary particle mass number A. However, the particle ratios at the detector are subject to shower-to-shower fluctuations, causing an inherent randomness and hence uncer-

tainty in the measurements. Nevertheless, the ability to measure both electronic and muonic components of the shower is beneficial for composition measurements, and has been used at KASCADE for a composition analysis with 5 primary mass groups i.e. a high resolution in composition (Antoni et al., 2005; Hörandel, 2008).

A solution is to measure Cherenkov light or fluorescence in the atmosphere. Both detection methods measure the faint trail of light produced by the particles as the shower front traverses the atmosphere. Cherenkov light is emitted because the particles travel faster than the speed of light in air, which is smaller than the vacuum speed of light by a factor $n \approx 1.0003$, n being the refractive index of air. This technique (also called non-imaging Cherenkov detection, to distinguish from Cherenkov telescopes for high-energy gamma rays) is used at the Tunka (Prosin et al., 2014) and Yakutsk (Ivanov et al., 2009) measurement sites.

Fluorescence light is generated from excitation of nitrogen molecules by the particle shower; on de-excitation, blue and UV-light is emitted. The intensity of this light is proportional to the energy deposit of the shower into the atmosphere. Moreover, it can be imaged, tracing the shower as it passes through the atmosphere. The point of maximum intensity defines $X_{\max}$. The accuracy of this method is about $20\,\mathrm{g/cm^2}$ (Aab et al., 2014a). Fluorescence detection is used e.g. at the Pierre Auger Observatory, Telescope Array (Matthews et al., 2011), and HiRes (Abu-Zayyad et al., 2000).

The downside of these optical methods is that they only work in dark, moonless nights free of clouds, thus limiting their operational duty cycle to below 15 %. Furthermore, atmospheric influences such as absorption by aerosols, and variations from the local weather need to be well understood and quantified. Detection of the radio signal emitted by the air shower particles, as discussed in the following sections, does not have this duty cycle limit; with the exception of thunderstorm conditions, radio detection can in principle have a near-100 % duty cycle.

1.5 Radio detection of cosmic rays

As the charged particles in the shower front pass through the atmosphere, they generate radio emission which is detectable in antennas on the ground as short pulses, on the order of 10 to 100 ns long. The pulses are strongest for frequencies

below 100 MHz. Therefore, for their detection one typically uses a frequency band from 30 to 80 MHz as this blocks out strong radio frequency interference (RFI) from transmitters below 30 and above 80 MHz. Moreover, below 30 MHz the emission from the Galaxy becomes too strong compared to the air shower signals.

The main mechanism that generates the radio emission is the *transverse current*, that arises from the deflection of electrons and positrons as they pass through the Earth's magnetic field. They are deflected in opposite directions, perpendicular to the incoming direction. This deflection of charges gives rise to a net current, transverse to the direction of motion of the air shower front. The current is also time-dependent, as the number of contributing particles rises up to the $X_{\max}$ level, and then falls again (Scholten et al., 2008; Kahn and Lerche, 1966).

A secondary mechanism is due to the negative *charge excess* that builds up in the air shower front, due to ionization. The free electrons are much more mobile than the heavier positive ions, hence an excess negative charge is found in the shower front (Askaryan, 1962). This makes up about 15 % of the electric field (Schellart et al., 2014; Aab et al., 2014b).

Radio detection of air showers was demonstrated first by Jelley et al. (1965). However, the available technology in those days did not allow for a detailed study of the radio emission pattern, as fast digital electronics was not yet available. Detections involved making photographs of oscilloscope traces, processing data was cumbersome and man-made interference could not be filtered out completely. Therefore, the technique was largely abandoned.

The revival came in the first decade of this century (see e.g. Huege, 2016; Huege et al., 2012; Schröder, 2017), when it was realized that a digital radio telescope consisting of many low-frequency antennas would solve these problems and allow for detailed and competitive studies (Falcke and Gorham, 2003; Falcke et al., 2005). The first successful demonstrations of radio detection were performed at LOPES, a LOFAR-prototype station in Karlsruhe, Germany which has operated from 2003 to 2013 (Falcke et al., 2005; Huege et al., 2012), and CODALEMA in Nançay, France (Ravel, 2012) which is still active today. The role of the measurements at the Low Frequency Array (LOFAR) is introduced in the next section (1.6), and they are the main subject of this thesis.

Other radio detection experiments are the Auger Engineering Radio Array (AERA)

Figure 1.3: Left: The LOFAR Superterp, the central circular core with 6 stations. Right: One of the LBA antennas.

(Schulz et al., 2015) at the Pierre Auger Observatory, TUNKA-Rex (Bezyazeekov et al., 2015) in Siberia, and experiments detecting GHz radiation such as CROME (Smida et al., 2014) and ANITA (Hoover et al., 2010).

1.6 The Low Frequency Array, LOFAR

The research in this thesis is focused on radio detection of air showers with the LOFAR radio telescope (van Haarlem et al., 2013), which has a core region located in the north of the Netherlands. It is a distributed radio telescope which as of 2019 consists of 51 stations, spread over several European countries, with a high density of antennas in the core region in the Netherlands. In one station, there are 96 low-band antennas (LBA) sensitive to the 10 to 80 MHz range, and 48 high-band antennas (HBA) observing from 110 to 240 MHz. For measuring cosmic rays, mainly the low-band antennas are used. With two Λ-shaped dipoles per antenna, covering two polarizations, these have an omnidirectional field of view, up to zenith angles of at least 60 degrees. Also, the radio signal of air showers is strongest below 100 MHz. The central region of LOFAR, which has 6 stations in a circular area of 320 m diameter (also known as *Superterp*, see Fig. 1.3), is the main set of antennas where air showers are observed. Inside this ring, an array of particle detectors called LOFAR Radboud Air Shower Array (LORA) has been set up (Thoudam et al., 2014, 2016). It comprises 20 scintillator detectors of 0.9 m^2 each, with photomultiplier tubes

to detect incoming particles. A detector triggers on exceeding a threshold of $4\,\sigma$ above the noise floor. From coincident triggers, a real-time reconstruction program determines basic parameters of the air shower, such as the incoming direction and core position on the ground, and the energy of the primary particle.

When LORA detects particles in coincidence in 13 or more detectors, a trigger is sent to LOFAR. The value is chosen to optimize the number of gathered showers while still having a good chance of a successful radio detection, and also not exceeding the data transfer limits set by the observatory. In practice, this amounts to detecting showers above about $3 \cdot 10^{16}\,\mathrm{eV}$, and not sending more than one trigger per hour.

When LOFAR receives a trigger, a buffer readout is done of 2 ms of data gathered with the LBA antennas around the time of arrival. This amounts to about 75 MB per LOFAR station. The data channel of each antenna dipole is equipped with a ring buffer (called Transient Buffer Board or TBB) which continuously stores the data as it comes in. It is just over 5 seconds long, allowing ample time for processing particle data, generating and sending the trigger. The use and readout of the TBBs operates independently, parallel to normal LOFAR observations. This is well suited for cosmic-ray studies, as they can gather data from long observation times while LOFAR is running astronomical observations. In practice however, the duty cycle of obtaining useful data is limited by the array setup chosen in the primary observations; in particular whether they observe using HBA or LBA antennas. Idle time at LOFAR is usually filled up with dedicated LBA observations for cosmic rays.

An example graph for a measured air shower, showing the pulse arrival times in the central ring and three additional stations, is given in the left panel of Fig. 1.4. There are in total 22 stations in a region of $2 \times 2.5\,\mathrm{km}$ around the central core. The right panel shows a pulse measured in one of the dipoles, which is the raw material for the analysis of air showers from their radio signal.

1.7 Measuring cosmic rays at LOFAR

The project to measure cosmic rays with LOFAR started in 2007. During the early years, LOFAR was still in its commissioning phase, hence data acquisition was still

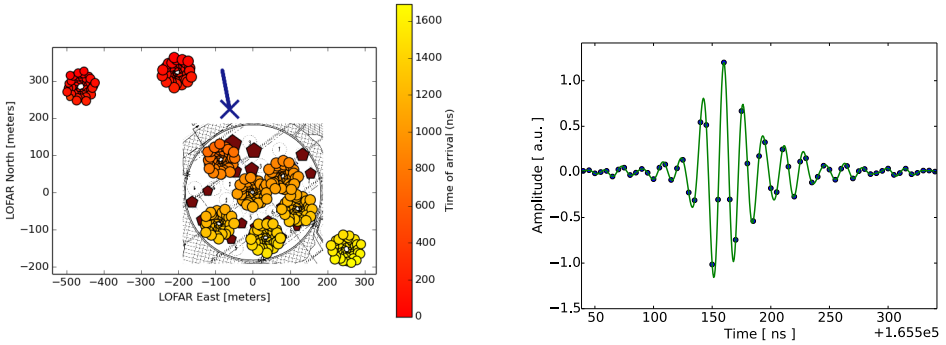


Figure 1.4: *Left*: An example of a cosmic-ray detection with a particularly strong signal, allowing detection in 9 stations. Colored circles represent LBA antennas, color-coded by the arrival time of the pulse. The pentagons show the positions of the LORA particle detectors, with the size of the symbols representing the number of particles detected. The blue line and cross indicates the reconstructed shower direction and core position, respectively. *Right*: An example of a pulse detected in one of the LOFAR LBA antennas. The dots represent the samples of the time series signal as measured from the antenna, and the connected curve shows the upsampled (interpolated) signal.

under development. The LORA particle detectors were installed and calibrated (Thoudam et al., 2014, 2016).

LOFAR has detected its first cosmic rays in June 2011, and is still measuring to this date. Meanwhile, a data analysis pipeline was developed. As standard astronomical software is not suitable for our purpose, the pipeline was custom made from scratch, for detection of cosmic-ray pulses on nanosecond timescales in raw time series data (Schellart et al., 2013). The pipeline performs spectral cleaning on the time series data, using an algorithm based on stability of relative phases between antennas. This is described in Chapter 3. This removes narrow-band transmitter signals from the data. After this, three of the four main properties of the pulse are measured: its intensity (or total energy in the pulse), its arrival time, and its polarization through the four Stokes parameters. The frequency spectrum of the pulses is still subject of study. An example pulse measured at an LBA antenna is

shown in the right panel of Fig. 1.4.

Measurements of the polarization have confirmed that the pulses are linearly polarized (Schellart et al., 2014). Their polarization angle approximately aligns with the $\vec{v} \times \vec{B}$ axis, where $\vec{v}$ is the incoming direction and $\vec{B}$ the magnetic field. This confirms that the transverse-current emission mechanism is dominant. The secondary emission mechanism due to negative charge excess is also seen, through a smaller component of the polarization vector pointing radially outward from the shower core. Later, a further refined analysis confirmed a component of circular polarization (Scholten et al., 2016), as is expected to arise from the interplay of the two main radio emission mechanisms.

We have developed a parametrization of the radio intensity footprint (or lateral distribution of intensity) of air showers (Nelles et al., 2015c,b), allowing a fast first estimate of shower core position, $X_{\max}$, and energy. We have developed a method which uses detailed air shower simulations which are fitted to measured showers (see next section), to estimate $X_{\max}$ to a resolution of $20\,\mathrm{g/cm^2}$ (Buitink et al., 2014). This has allowed the first cosmic-ray composition measurements at LOFAR (Buitink et al., 2016). The analysis and results in Chapter 5 are an extension of these.

From the arrival times, the incoming direction is estimated using a plane wave assumption; in Chapter 2, it is shown that the shape of the incoming wavefront is in fact hyperbolic, and that measuring the hyperboloid shape helps to improve the reconstruction of the incoming direction.

Despite their limited field of view, we have been able to use the high-band antennas to observe radio footprints shaped like a Cherenkov ring, in the 110 to 190 MHz range (Nelles et al., 2015d).

During thunderstorms, the radio emission of air showers changes, due to the significant electric fields in the atmosphere. This is explained in Trinh et al. (2016). The effect can be used, to measure atmospheric electric fields using air showers (Schellart et al., 2015; Trinh et al., 2017).

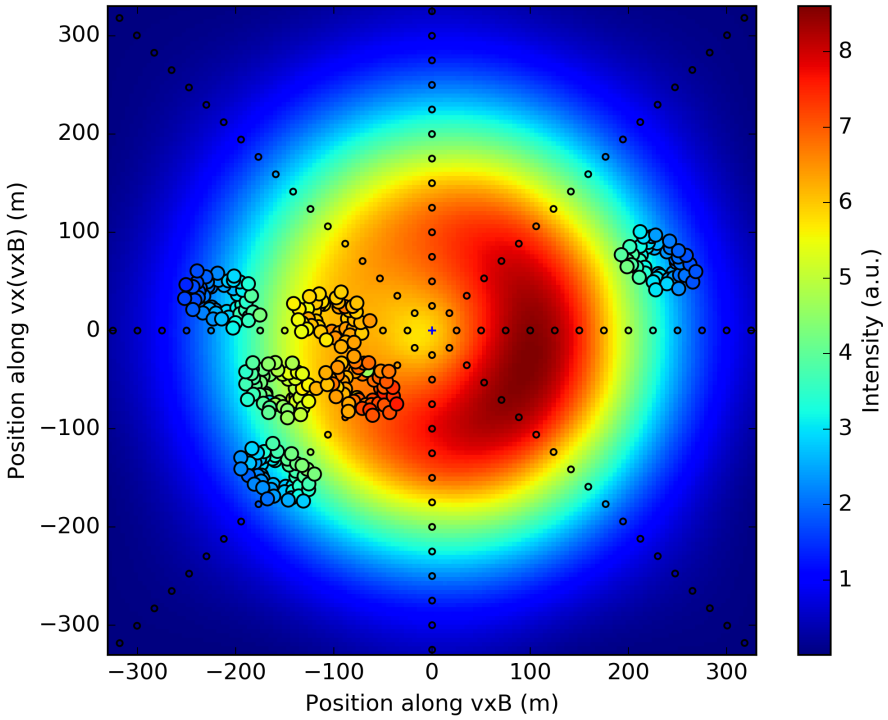


Figure 1.5: Radio intensity footprint of a simulated air shower, fitted to an air shower measured with LOFAR. Color represents intensity, increasing from blue to red. The small circles show the positions on which the electric field was simulated, while the bigger circles show the LOFAR antennas. Positions have been projected onto the shower plane, perpendicular to the incoming direction.

1.8 Simulations of air showers and comparison to measurements

To estimate $X_{\max}$ and infer the composition of a collection of air showers, the radio method relies on simulations of air showers. The air shower simulation code Corsika (Heck et al., 1998) simulates the particle interactions in the atmosphere, tracking the air shower propagation and its secondary particles. This is done using a Monte Carlo method, as the particle interactions, their decay, and the type and direction of the newly produced secondary particles are inherently random. Therefore, when simulating an ensemble of air showers with the same primary particle and energy, the random variations found among the simulated showers represent the naturally occurring variations from shower to shower.

The CoREAS (Huege et al., 2013) plugin to Corsika follows the simulated particles in the air shower, and computes for each particle its contribution to the radio emission detected at given positions on the ground. This is done using the 'end-point formalism' (James et al., 2011), in which a particle's track is split up in sections. One computes the radiation produced by 'accelerating' a particle at the beginning of a section from zero to its given velocity, and decelerating it again to zero at the end of the section. In curved trajectories, subsequent sections are not fully aligned, and the net acceleration leads to radiation. This scheme has been shown to accurately approximate the actual radiation, while allowing for a practical implementation in simulation codes.

It is noteworthy that the radiation that reaches the ground is simulated completely from first principles, i.e. Maxwell's equations. Thus, the simulation is independent of specific emission mechanisms or parametrizations of particle distributions. It takes as input the Earth's magnetic field vector at the detector site, and the index of refraction as a function of altitude in the atmosphere. This way, the emission caused by curved trajectories in the magnetic field, and the Cherenkov-like effects caused by particles propagating faster than the speed of light in air, are automatically taken into account. In Chapter 4, the effect of variations in refractive index with local weather conditions is investigated. It is shown that for precision work, these variations are significant and should not be neglected. This has led to an improvement in CoREAS, which instead of relying on average profiles, can now

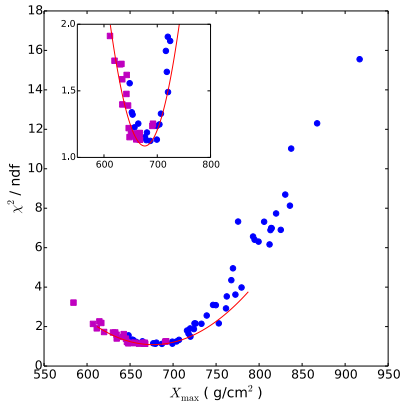


Figure 1.6: Example plot of fit χ^2 for an ensemble of simulated air showers, when fitted to a footprint measured with LOFAR. The simulations were done with iron primary particles (purple squares) and protons (blue dots). Figure taken from Buitink et al. (2014).

take local atmospheric profiles as input, including the refractive index.

Other simulation codes include ZHAires (Alvarez-Muñiz et al., 2012), which also simulates individual particles and produces similar results to CoREAS. The SELFAS2 (Marin and Revenu, 2012) code is based on a more macroscopic approach, where charge distributions are used instead of individual particles. This is expected to be less accurate in general, but also more easily allows to study the different emission mechanisms separately.

An example of a simulated lateral distribution of the radio intensity, also simply called the 'radio footprint', is shown in Fig. 1.5. It is matched to LOFAR measurements of an example air shower by least-squares fitting, with the core position and a scaling factor as free parameters. The colors of the circles represent the measured intensity in each antenna. They mostly blend into the colors of the simulated footprint, indicating a good fit. For measuring $X_{\max}$, an ensemble of 40 air showers is simulated (see Buitink et al., 2014), given the measured arrival direction and energy estimate. The simulation shown in Fig. 1.5 is the best-fitting simulation of the ensemble for this measured air shower. See also the book cover for a simulated footprint on the ground, plus corresponding LOFAR measurements.

In this ensemble, one uses 25 proton primaries and 15 iron nuclei, to span the entire range of possible $X_{\max}$ levels. Each of the showers is fitted to the measurements this way, including the simulated particle content which is fitted to the measured number of particles at LORA. This gives a χ^2-value for each shower. When these are ordered by $X_{\max}$, their χ^2-values have a minimum, and to lowest order, follow a parabola around the minimum (see Fig. 1.6). The scatter around this parabola represents shower-to-shower fluctuations; fitting a parabola yields the best estimate for $X_{\max}$. This technique is used in the composition analysis in Buitink et al. (2016) as well as in the refractivity study in Chapter 4 and in the refined composition analysis in Chapter 5.

1.9 This thesis

Over the last years, since its first cosmic-ray detection in 2011, LOFAR has proven to be a valuable and high-precision experiment for detecting and measuring air showers. Its dense array features an order of magnitude more antennas than typical for dedicated air shower arrays. This shows the synergy between astronomy and particle physics, making LOFAR an *astroparticle physics* instrument as well.

In Chapter 2, I demonstrate high-precision measurements of the shape of the radio shower front as it arrives on the ground. Reaching nanosecond timing resolution, LOFAR was the first experiment to achieve this for individual air showers. The opening angle of the wavefront was found to deviate from a plane wave by about one degree, and spherical, conical, and hyperboloid shapes have been fitted. The hyperboloid describes all measured shapes, although a subset of air showers can also be described by a cone or sphere over the extent of the array. The cone and sphere are limiting cases of a hyperboloid, for extreme choices of its two parameters.

The achieved precision in the wavefront measurements have been made possible only through accurate (sub-nanosecond) timing calibration between LOFAR stations, as well as spectral cleaning of the measured time series. As LOFAR, as an astronomical telescope, typically deals with hardware-preprocessed data, which is correlated and integrated over timescales of hours, the customary calibration algorithms cannot be used on short, raw time series spanning no more than a few

milliseconds. Thus, dedicated algorithms were required to perform this for the purpose of air shower measurements.

In Chapter 3, I present an algorithm for timing calibration, based on relative phases between antennas. Measuring the signal from a stationary reference transmitter at a known position, which in this case is a commercial FM radio station, gives a stable phase offset between each pair of LOFAR antennas, corresponding to the geometric delay of signal propagation. Deviations from this value are then calibrated out.

Realizing that when no transmitter signal is present, the relative phases are uniform-random instead (the background comes from a large number of sources in the Galaxy), I found that the same algorithm would also perform spectral cleaning. For this, one measures not only the relative phase between antenna pairs, but also the phase variance over roughly 50 data blocks in the 2 ms-long measurement. Low variance at a given frequency means a transmitter is present, and this frequency is flagged as RFI. The use of phases is a new approach to RFI cleaning, and performs about equally well as amplitude-based algorithms, while being easier to implement as the antenna's frequency characteristics need not be known.

The cosmic-ray mass composition results for the energy range 10^{17} to 10^{18} eV in Buitink et al. (2016) have achieved a high precision in $X_{\max}$ of $20\,\mathrm{g/cm^2}$ per shower, such that in later, higher-statistics studies, the remaining systematic effects become important as well. One of the most important systematics comes from the index of refraction of the air in the atmosphere, as I show in my analysis in Chapter 4. It followed that variations due to local weather cannot be neglected, and this has led to an improvement in the CoREAS simulations, which now use as input the atmospheric parameters at the time of measurement, which are taken from the global database GDAS (National Oceanic and Atmospheric Administration, 2010).

With this improvement, and a larger dataset, I have revised the composition measurement. This is the subject of Chapter 5. By determining $X_{\max}$ of each shower using a radio-only fit method, using simulations that take the local atmospheric parameters into account, I arrive at a systematic uncertainty on $X_{\max}$ of $7\,\mathrm{g/cm^2}$. A radio-only measurement of the primary energy, based on an improved calibration of the LOFAR antennas (Mulrey et al., 2019), allows for a precision of 10 %, at a systematic uncertainty of 27 %. Using improved criteria for selecting a bias-free

sample, I have analyzed a dataset of 196 measured showers. A maximum-likelihood based statistical analysis shows a proton fraction that is about 18 % at best fit, with an upper bound of 43 %, at 99 % confidence level, over all three important hadronic interaction models considered. The largest fraction of particles is in the intermediate-mass range, centered around carbon, nitrogen and oxygen. Depending on the choice of hadronic interaction model, the iron fraction (heavy particles) is between 5 and 38 % at best fit.

Chapter 2

The shape of the radio wavefront of extensive air showers as measured with LOFAR

A. Corstanje et al. (LOFAR Collaboration)

Astroparticle Physics 61, 22-31 (2015)

Abstract

Extensive air showers, induced by high energy cosmic rays impinging on the Earth's atmosphere, produce radio emission that is measured with the LOFAR radio telescope. As the emission comes from a finite distance of a few kilometers, the incident wavefront is non-planar. A spherical, conical or hyperbolic shape of the wavefront has been proposed, but measurements of individual air showers have been inconclusive so far. For

a selected high-quality sample of 161 measured extensive air showers, we have reconstructed the wavefront by measuring pulse arrival times to sub-nanosecond precision in 200 to 350 individual antennas. For each measured air shower, we have fitted a conical, spherical, and hyperboloid shape to the arrival times. The fit quality and a likelihood analysis show that a hyperboloid is the best parametrization. Using a non-planar wavefront shape gives an improved angular resolution, when reconstructing the shower arrival direction. Furthermore, a dependence of the wavefront shape on the shower geometry can be seen. This suggests that it will be possible to use a wavefront shape analysis to get an additional handle on the atmospheric depth of the shower maximum, which is sensitive to the mass of the primary particle.

2.1 Introduction

A high-energy cosmic ray that enters the atmosphere of the Earth will interact with a nucleus of an atmospheric molecule. This interaction produces secondary particles, which in turn interact, thereby creating a cascade of particles: an *extensive air shower*. The origin of these cosmic rays and their mass composition are not fully known.

Due to the high incident energy of the cosmic ray, the bulk of the secondary particles propagate downward with a high gamma factor. As this air shower passes through the atmosphere and the Earth's magnetic field, it emits radiation, which can be measured by antennas on the ground in a broad range of radio frequencies (MHz - GHz) (Allan and Jones, 1966; Jelley et al., 1965; Falcke et al., 2005). For a review of recent developments in the field see Huege (2013). The measured radiation is the result of several emission processes (Huege et al., 2013), and is further influenced by the propagation of the radiation in the atmosphere with non-unity index of refraction (Werner et al., 2012). Dominant in the frequency range considered in this study is the interaction in the geomagnetic field (Kahn and Lerche, 1966; Allan, 1971; Falcke et al., 2005; Ardouin et al., 2009). An overview of the current understanding of the detailed emission mechanisms can be found in Huege (2012).

The radio signal reaches the ground as a coherent broadband pulse, with a duration on the order of 10 to 100 ns (depending on the position in the air shower geometry). As the radio emission originates effectively from a few kilometers in altitude, the incident wavefront as measured on the ground is non-planar. Geometrical considerations suggest that the amount of curvature and the shape of the wavefront depend on the height of the emission region, suggesting a relation to the depth of shower maximum, $X_{\max}$. The depth of shower maximum is related to the primary particle type.

Assuming a point source would result in a spherical wavefront shape, which is used for analysis of LOPES data (Nigl et al., 2008). It is argued in Schröder et al. (2011) that the actual shape of the wavefront is not spherical, but rather conical, as the emission is not point-like but stretched along the shower axis. In a recent further refinement of this study, based on CoREAS simulations, evidence is found for a hyperbolic wavefront shape (spherical near the shower axis, and conical further out) (Apel et al., 2014). Hints for this shape are also found in the air shower dataset collected by the LOPES experiment (Schröder et al., 2012a). However, due to high ambient noise levels, the timing precision of these measurements did not allow for a distinction between spherical, hyperbolical and conical shapes on a shower-by-shower basis, and only statistically was a hyperbolic wavefront shape favored.

We use the LOFAR radio telescope (van Haarlem et al., 2013) to measure radio emission from air showers, in order to measure wavefront shapes for individual showers. LOFAR consists of an array of two types of antennas: the low-band antennas (LBA) sensitive to frequencies in a bandwidth of $10 - 90$ MHz, and the high-band antennas (HBA) operating in the $110 - 240$ MHz range. While air showers have been measured in both frequency ranges (Schellart et al., 2013; Nelles et al., 2015d), this study only uses data gathered with the $10 - 90$ MHz low-band antennas. A combination of analog and digital filters limits the effective bandwidth to $30 - 80$ MHz which has the least amount of radio frequency interference. For detecting cosmic rays we use the (most densely instrumented) inner region of LOFAR, the layout of which is depicted in Fig. 2.1. LOFAR is equipped with ring buffers (called Transient Buffer Boards) that can store the raw-voltage signals of each antenna for up to 5 seconds. These are used for cosmic-ray observations as described in Schellart et al. (2013).

Inside the inner core of LOFAR, which is a circular area of 320 m diameter, an

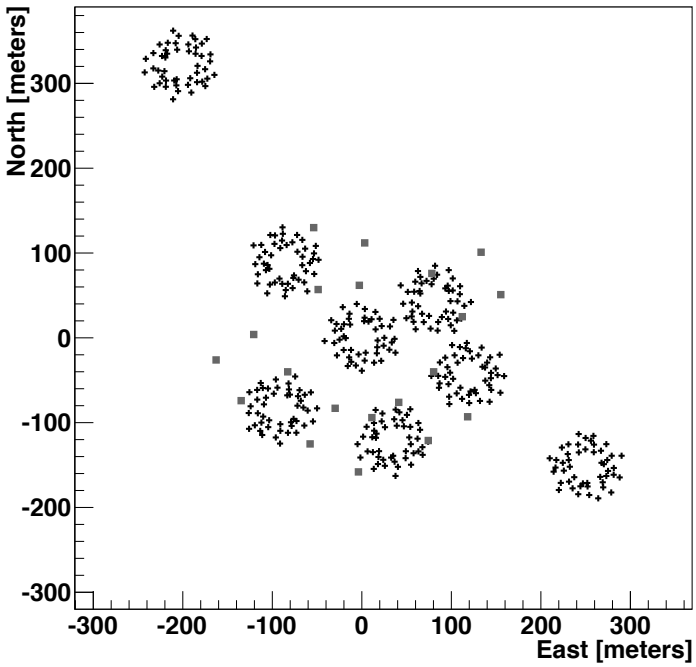


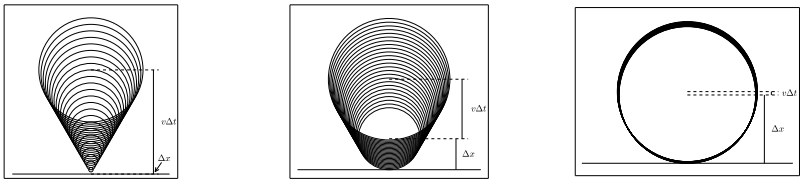
Figure 2.1: Layout of the innermost 8 stations of LOFAR. For each station the outer ring of low band radio antennas (black plus symbols), used for the analysis in this paper, are depicted. Located with the innermost six stations are the particle detectors (grey squares) used to trigger on extensive air showers.

array of 20 scintillator detectors (LORA) has been set up (Thoudam et al., 2014). This air shower array is used to trigger a read-out of the Transient Buffer Boards at the moment an air shower is detected. The buffer boards provide a raw voltage time series for every antenna in a LOFAR station (a group of typically 96 LBA plus 48 HBA antennas that are processed together in interferometric measurements), in which we identify and analyze the radio pulse from an air shower. Analysis of the particle detections delivers basic air shower parameters such as the estimated position of the shower axis, energy, and arrival direction.

The high density of antennas of LOFAR, together with a high timing resolution (200 MHz sampling rate) are especially favorable for measuring the wavefront shape.

2.2 Simplified model for the wavefront shape

Inspection of the pulse arrival times in our datasets (as explained in the following sections) shows that while the shape at larger distances from the shower axis might be described by a conical wavefront, in many measured air showers there is significant curvature near the shower axis. A natural choice for a function of two parameters that describes this behavior is a hyperbola. In Fig. 2.2 a toy model is sketched, where the wavefront is formed by assuming the emission to be generated by a point source. At any given time, during the emission generation, the source generates a spherical wavefront that expands at the local speed of light c/n, where n is the local index of refraction. The point source is moving at a velocity $v > c/n$, and emits radiation for a limited amount of time Δt. In real extensive air showers this corresponds to the duration in which the bulk of the radiation is generated. The radiation is measured by an observer at a distance Δx from the point where the emission stopped. When this distance is small, $\Delta x/(v\,\Delta t) < 1$, (panel A) the combined wavefront shape is approximately conical. Even so, unless the distance to the last emission point is zero, a small curvature is visible near the shower axis, the radius of curvature corresponding to the distance. When viewed at intermediate distances, $\Delta x/(v\,\Delta t) \approx 1$, (panel B), the opening angle of the conical part increases and the curved part near the shower axis extends a bit further outward. This shape is closely approximated by a hyperbola. Only when the distance to the last emission point is very large compared to the duration of the emission times the local speed of light,



(a) Small (b) Intermediate (c) Large

Figure 2.2: Toy model motivating a hyperbolic wavefront shape. A point source moves vertically at a velocity $v > c/n$ and emits for a limited amount of time. The solid horizontal line represents the ground plane. The generated wavefront is observed as conical (panel A) by an observer at small distances to the point where the source stops emitting. Observers at intermediate distances see a hyperbolic wavefront shape (panel B). For observers at larger distances the observed wavefront shape is closer to a sphere (panel C).

$\Delta x/(v\,\Delta t) \gg 1$, is the wavefront shape approximately spherical. In this simplified picture, with constant but non-unity index of refraction, the wavefront shape is thus hyperbolic for most observer distances. We expect the general characteristics of this simplified model to hold even for a realistic atmosphere where the refractive index changes with height as the main criterion $n > 1$ still holds true.

Motivated by this toy model we therefore compare three parametrizations of the wavefront shape: a sphere, a cone and a hyperboloid, and evaluate the quality of the fits to the LOFAR measurements.

2.3 Measurements

For this analysis we have used air-shower measurements with LOFAR accumulated between June 2011 and November 2013. These consist of 2 to 5 milliseconds of raw voltage time series for every antenna of the LOFAR core stations; we identify the air shower's radio pulse in every individual trace, and measure its strength and arrival time.

In order to have a dense, high-quality sampling of the radio wavefront, and a substantial distance range of more than $\sim 150\,\mathrm{m}$, we require an air shower to be detected in at least four LOFAR core stations. Furthermore, the highest quality data are obtained with the outer ring of low-band antennas, as its more sparse layout gives a more even distribution of measurements over the area of interest. Therefore, the sample is restricted to this subset. This leaves a total of 165 measured air showers. Of these 165, three fail calibration of time differences between stations due to corrupted data (see Sect. 2.3.2) and one is unreliable due to thunderstorm conditions (see Sect. 2.4.6). This leaves a total of 161 high quality air shower measurements for this analysis.

All measured air showers are processed by the standard cosmic-ray reconstruction software as described in Schellart et al. (2013).

2.3.1 Pulse arrival times & uncertainties

The arrival time of the radio pulse in each dipole is determined using the raw-voltage traces. We define the arrival time as the time of the pulse maximum in the amplitude (or Hilbert) envelope of the analytic signal $A(t)$:

$$A(t) = \sqrt{x^2(t) + \hat{x}^2(t)}, \tag{2.1}$$

where $\hat{x}(t)$ is the Hilbert transform of the voltage-trace signal $x(t)$, upsampled by a factor of 32. The Hilbert transform is defined by

$$\mathcal{F}\left[\hat{x}(t)\right](\omega) = -i\ \mathrm{sgn}(\omega)\ \mathcal{F}\left[x(t)\right](\omega), \tag{2.2}$$

where $\mathcal{F}$ is the Fourier transform. This allows for arrival time measurements at a much higher time resolution than suggested by the 200 MHz sampling rate (5 ns sampling period).

The attainable timing precision varies with the signal-to-noise ratio (S/N). Uncertainties in the arrival time are assigned independently to each datapoint using the measured S/N in amplitude following

$$\sigma_{t_{\max}} = \frac{12.65}{S/N}\ \mathrm{ns}. \tag{2.3}$$

A similar relation was found in Schröder et al. (2012b). The one used here is derived from the data for each antenna, using a procedure as follows.

Uncertainties on the timing arise from distortions of the pulse shape due to fluctuations in the background noise. These uncertainties on the timing can be determined from the amplitude data for each antenna. To quantify them, we first select a noise block outside of the pulse region for each antenna and calculate the root-mean-square noise level N. We also calculate the signal-to-noise ratio S/N of the pulse maximum (in amplitude). Subsequently we add this noise to the data containing the pulse, and the pulse arrival time is calculated using the procedure described above. This procedure is repeated 10 times, where at each time the noise block is shifted by 100 ns. This gives 10 measurements of the pulse arrival time. The standard deviation of this set for each antenna is a measure of the uncertainty on the determination of the pulse arrival time. However, because this procedure effectively reduces S/N by a factor of $\sqrt{2}$, pulses with a low S/N are no longer correctly identified. Therefore, instead of assigning this uncertainty to the datapoint directly, we estimate the uncertainties as a function of S/N by processing data for all antennas in the full air shower dataset. Uncertainty data points are binned with respect to S/N. The bin size is set to 1.0 below $S/N = 50$ and to 10.0 above to ensure a sufficient number of points per bin. To prevent outliers (due to accidental spike selection in some antennas) from heavily influencing the results, the median and uncertainty on the median are calculated for any given S/N bin. The result can be seen in Fig. 2.3.

One can see that the timing uncertainty is inversely proportional to the signal-to-noise ratio. Fitting this relation gives the proportionality factor in Eq. 2.3, and we use this to assign an uncertainty to the arrival time of the pulse maximum for each antenna depending on its measured S/N.

2.3.2 Time differences between stations

For time calibration of individual antennas within one LOFAR station, we use standard LOFAR calibration tables as described in Schellart et al. (2013). Since all LOFAR core stations share a single clock these calibration solutions are stable over time. However, before October 10th 2012 this common clock was only available for the innermost region (consisting of 6 stations). Every other core station had its own clock, synchronized by GPS[1]. Drifts of these clocks with respect to each other

[1]Global Positioning System

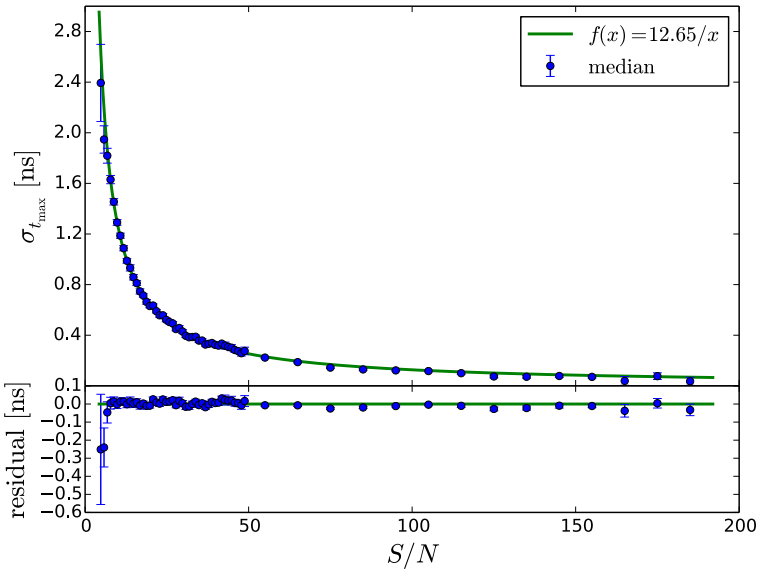


Figure 2.3: Uncertainties on the determination of the pulse arrival time as a function of signal to noise of the pulse amplitude. For each S/N bin (of width 1.0 for $S/N < 50$ and width 10.0 for $S/N \geq 50$) the circular dots (blue) give the median value of the uncertainties. Error bars represent the standard error on the median in each bin. The solid line (green) represents the fitted relation; the lower panel shows the residual to the fit.

were on the order of 10 ns which is large compared to the other timing uncertainties. Datasets taken before this date therefore require a more involved calibration procedure described below.

Every air shower raw voltage trace is only 2 ms long, and all measured air showers are scattered over a 2-year timespan. Therefore, using dedicated calibration observations is not feasible, as these are typically planned in advance and can take minutes to hours. Instead, we make use of radio transmitter signals present in every dataset. These transmitters emit continuous waves, measured at each antenna with a different phase. We use the Fast Fourier Transform to calculate the phase per antenna and frequency channel. The phase differences between antenna pairs can be used (directly) to monitor and correct for deviations from a trusted timing calibration; this technique was originally developed by the LOPES experiment as demonstrated in Schröder et al. (2010).

In addition to this, we make use of the (known) position of the transmitter in order to predict the relative phases at every antenna pair (accounting for the geometric delays). The difference in measured versus expected (relative) phase, averaged over each LOFAR station, yields the inter-station clock calibration. It was found that higher-frequency signals can be measured with better phase stability; the public radio transmitters at 88 to 94 MHz are suitable for this purpose.

We use the (strongest) 88.0 MHz transmitter to fix the station's clock offsets modulo its period, 11.364 ns. The remaining ambiguity is resolved using trial and error in the wavefront analysis, incrementally adding outer-core stations to the (already calibrated) inner core data. As 11 ns is large compared to the expected wavefront arrival time delays between stations, there is only one best-fitting solution.

Differences in filter characteristics or propagation effects between antennas in a station are expected to average out over an entire station, leading to a calibration of the station clock offsets to about 0.3 ns precision.

2.3.3 Shower parameters

An independent reconstruction of the shower is performed based on the detected particle density in each scintillator detector, as described in Thoudam et al. (2014). This yields the direction and location of the shower axis, as well as an energy estimate. However, these reconstructed values are only reliable for a restricted pa-

rameter space; for example the shower axis should fall inside the instrumented area. In order to retain a substantial set of showers for this analysis, these cuts are not applied to the set of radio measurements. Therefore, we do not have reliably reconstructed shower axis locations for all measured air showers and the core positions are only used as initial estimates that are optimized later (see Sect. 2.4.4). The reconstructed direction inferred from particle densities is however independent of the quality of other reconstructed quantities.

2.4 Reconstructing the wavefront shape

From the arrival time of the pulse in different radio antennas, and the information from the particle detector array, we find the shape of the wavefront using the following procedure.

2.4.1 Plane-wave approximation and curvature

We infer the general direction of the incoming pulse by obtaining the best-fitting plane wave solution to the arrival times of the radio pulse:

$$c\,t_i = Ax_i + By_i + C, \tag{2.4}$$

where t_i and (x_i, y_i) are the arrival times and antenna positions respectively. This holds for an antenna array for which all antennas lie in the same plane at constant z, which is true for LOFAR's inner core region. The fitted parameters A and B yield the azimuth angle ϕ and zenith angle θ, from the relations

$$\theta = \arcsin\left(\sqrt{A^2 + B^2}\right), \tag{2.5}$$

$$\phi = \frac{\pi}{2} - \arctan(B/A), \tag{2.6}$$

where the angle from the arctangent is taken in the appropriate quadrant. The global offset C is not used here. We can subtract the arrival times of the best-fitting plane wave from the measured times. This gives the curvature of the wavefront with respect to the array barycenter, defined as the average of the (x, y) coordinates for antennas with data.

2.4.2 Shower plane geometry

Given the shower axis position and direction, we can make a one-dimensional plot of the wavefront as a function of the distance to the shower axis. This assumes axial symmetry of the wavefront. In order to do this, all antenna positions are projected into the shower plane (defined by the shower axis as its normal vector), see Fig. 2.4. A one-dimensional function describing the wavefront curvature can then be fitted to the arrival times as a function of distance to the shower axis.

The projection into the shower plane is (by its definition) performed along lines parallel to the shower axis. This is an approximation, as the true wavefront is not planar, but has a small deviation angle α with respect to the shower plane. The angle may depend on the distance to the shower axis but can generally, for large distances be taken close to $1\,^\circ$. Projecting perpendicular to the true wavefront rather than the shower plane would give, to first order in α, a correction to the projected distance to the shower axis r, of $\Delta r/r = \tan(\alpha)\tan(\theta)$. For zenith angles below $45\,^\circ$, and for the longest distances in our dataset of about 500 m, this could introduce a timing uncertainty (scatter) of at most 0.5 ns at the largest distances (or 0.3 ns for typical zenith angles around 30 degrees). This is comparable to the calibration uncertainty between LOFAR station's clocks (see Sect. 2.3.2). A possible bias from an asymmetric antenna layout would be no larger than 2% on the estimated angle α and would not change the best-fitted shape.

Therefore, for the purpose of this analysis we apply the shower-plane projection, noting that in a detailed comparison with simulations, and for very large and/or inclined air showers, a two-dimensional fitting procedure may be favored.

2.4.3 Fitting the wavefront shape

Various wavefront shapes have been proposed; we test a conical and spherical shape, such as argued for in Schröder et al. (2011). We also test a hyperboloid; this is a natural function with 2 parameters that combines a curved shape for small distances, and a conical shape for large distances.

The fit functions, for the arrival time differences with respect to a plane wave as a function of distance to the shower axis, are those for a line (cone), a circle (sphere)

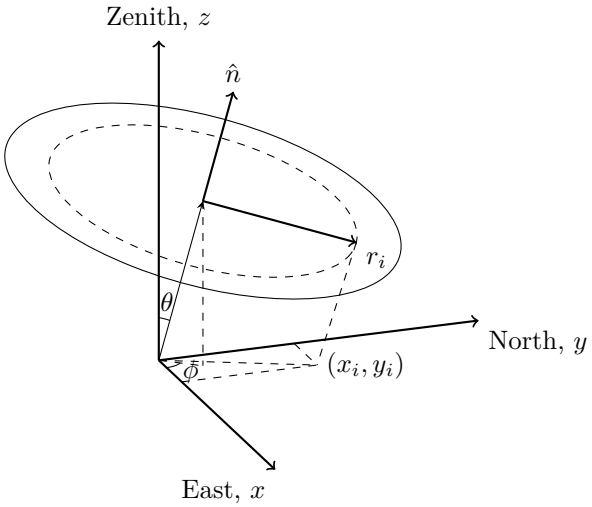


Figure 2.4: Geometry of the shower plane for a shower arriving from azimuth ϕ and zenith angle θ. The direction of the shower plane is defined by its normal vector $\hat{n}$. All antenna positions (x_i, y_i) are projected onto this plane giving r_i, the distance to the shower axis.

and a hyperbola (hyperboloid) respectively:

$$c\,t_{\mathrm{con}}(r) = s\,r, \tag{2.7}$$

$$c\,t_{\mathrm{sph}}(r) = \sqrt{R^2 + r^2} - R, \tag{2.8}$$

$$c\,t_{\mathrm{hyp}}(r) = -a + \sqrt{a^2 + b^2\,r^2}, \tag{2.9}$$

where s is the cone slope, R the radius of the sphere, and a and b are the parameters of the hyperboloid. These three functions are fitted in a standard non-linear least squares approach; the shower core x and y positions, needed to get the distance r, are used as free parameters in the fit, as explained further in the next section. We keep for each fit type the best fitting parameters as well as the fit quality, as measured by the unreduced χ^2 value

$$\chi^2_{\mathrm{type}} = \sum_{i=0}^{N_{\mathrm{antennas}}} \frac{(t_{\mathrm{type}}(r_i) - t_i)^2}{\sigma_i^2}, \tag{2.10}$$

where t_i is the arrival time of the pulse maximum at antenna i corrected for the best fitting plane wave solution and σ_i the corresponding uncertainty calculated using equation 2.3.

2.4.4 Considerations for fit stability

As the arrival time differences from a plane wave solution, and thus the shape of the wavefront, are sensitive to the direction and location of the shower axis, we include these as free parameters in the fitting procedure. If the core position would be well known, e.g. from signal amplitudes and/or comparison with simulations, each fit would have fewer degrees of freedom. Therefore typically, comparing the fit qualities of each shape, we find a lower bound to the differences with respect to the best fit.

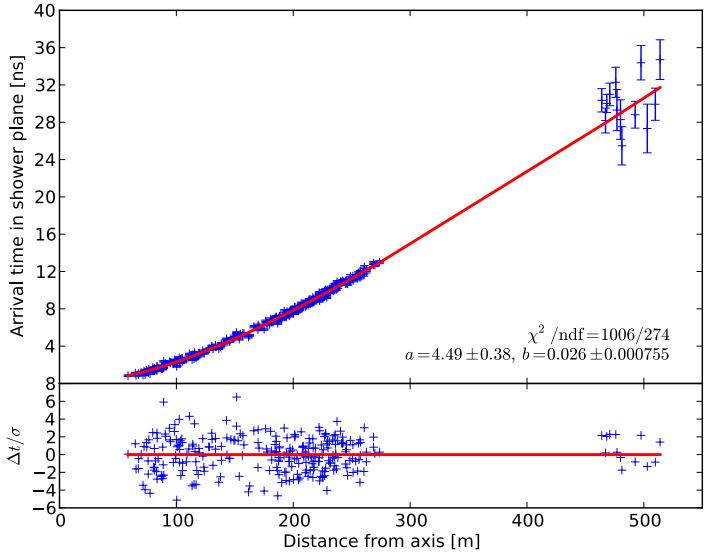
To prevent the fit from becoming unstable or finding only local minima we choose a nested approach. For every trial of the shower axis location, we optimize the direction; for every trial of the direction, we calculate the best-fit curve parameters using a nonlinear least-squares solver. Furthermore, to prevent the shower axis location search from getting stuck in a local minimum it is first optimized on a 500 m by 500 m grid in steps of 100 m and only in later iterations optimized further using a Nelder-Mead simplex optimization, starting from the optimal grid position.

2.4.5 Including particle detector information

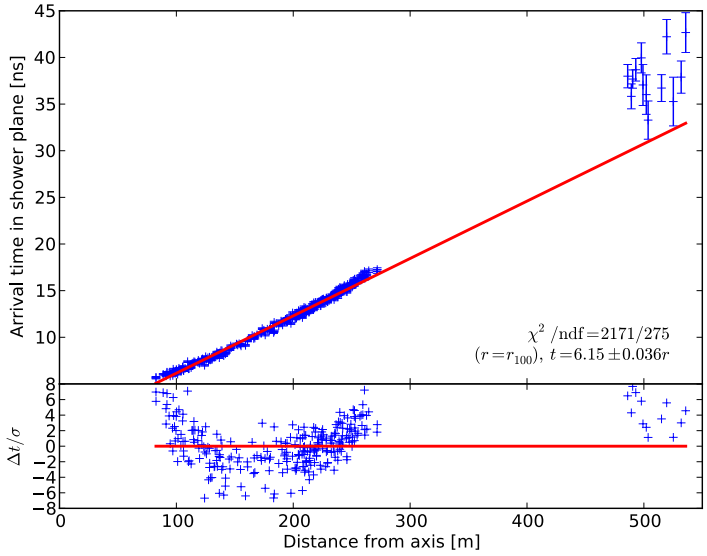
When optimizing the shower axis location, it might happen that the position is not well constrained due to the geometric distribution of the measurements. Furthermore, fitting a non-correct wavefront may also lead to an unphysical shower axis location. Typically this takes the form of the shower axis location moving too far away from the measured barycenter. The data from the particle detectors provide a further constraint on the shower axis location. The lateral distribution of the signal (number of particles as a function of distance to the shower axis) is well described by a Nishimura-Kamata-Greisen function (NKG) (Kamata and Nishimura, 1958; Greisen, 1960) and will restrict the position of the shower axis. Therefore, in the fit procedure we also re-fit the particle detector data using an NKG-function and add the (unreduced) χ^2 of this fit to the total χ^2 to minimize. This has a small influence when the shower axis location is within reasonable distances (due to the much larger number of radio measurements) but starts to dominate when the shower axis location moves to a position not supported by particle data. Note that the stored χ^2_{type} of the optimal curve fit does *not* include the particle fit χ^2.

2.4.6 Thunderstorm observations

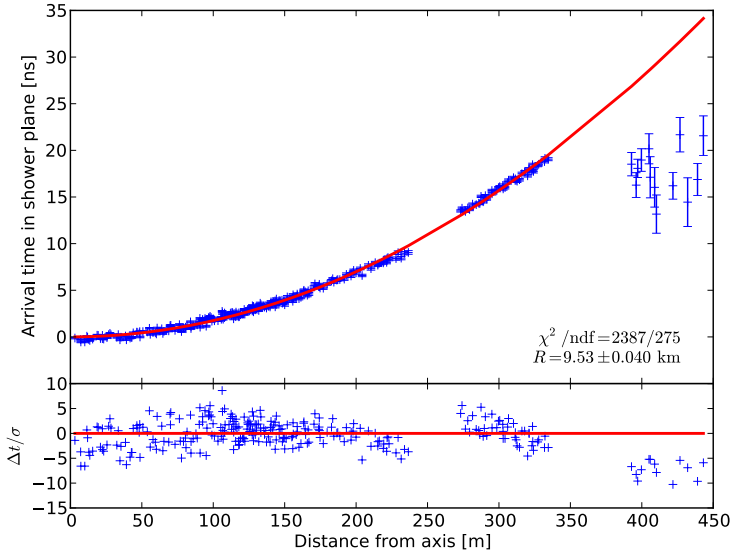
It has been reported that measured radio signals of air shower are amplified during thunderstorm conditions, which is attributed to the acceleration of electrons and positron in the electric fields (Buitink et al., 2007). In order to avoid a bias in the wavefront analysis, we have excluded one air shower that was measured during thunderstorm conditions. As the current dataset was taken without a local electric-field meter, the definition of thunderstorm conditions is based on lightning detections as provided by the Royal Dutch Meteorological Institute (Wessels, 1998). This cut does not account for local electric fields that do not manifest in lightning strikes. A local electric field meter is currently being installed at LOFAR. Future analyses will investigate the effects of the local electric field on air shower radio emission in more detail.



(A) Hyperbolic fit



(B) Conical fit



(c) Spherical fit

Figure 2.5: The arrival time differences from a plane wave as a function of distance to the shower axis with the best fitting shape solutions. A hyperbolic (top), conical (middle) and spherical (bottom) fit has been applied, respectively. Each plot shows the arrival times as a function of the distance to the shower axis (top panel) and deviations from the best fit scaled to the uncertainty for each datapoint (bottom panel). Note that the shower core position is a free parameter in each fit, therefore the positions of the data points on the x-axis differ between fits, as is in particular evident for the spherical fit.

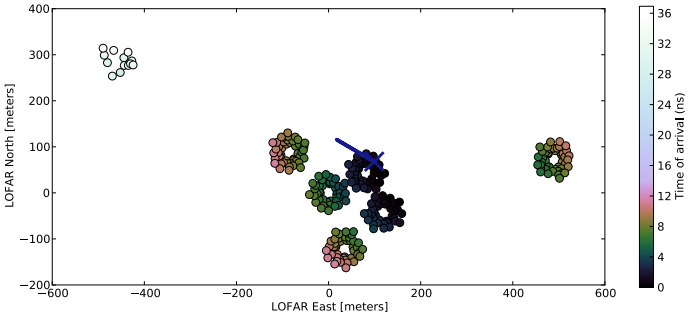


Figure 2.6: Relative arrival times for an example air shower measured with the LOFAR low band antennas. Circles indicate LBA antenna positions and their color corresponds to the measured pulse delay with respect to the best fitting plane wave solution. The shower axis (as determined from the particle detector data) is indicated by the blue line corresponding to the azimuthal arrival direction and cross where it intersects the ground.

2.5 Results

An example shower is shown in Fig. 2.6. This plot shows the layout of the LOFAR low-band antennas in the inner-core region. The colors show deviations from the best-fitting plane-wave solution, increasing outward from the center of the array.

2.5.1 Wavefront shape

The resulting best fitting wavefront shapes are given in Fig. 2.5a, 2.5b and 2.5c for a hyperbolic, conical and spherical wavefront, respectively.

The wavefront shape of this air shower is best fitted by a hyperbola due to significant curvature near the shower axis. The shower core position, left as free parameters in the fitting procedure, is significantly different for the three fits, as shown in Fig. 2.7. Fig. 2.8 shows the χ^2/ndf values obtained for all showers. From these distributions it is not immediately evident which wavefront shape (if any) is favored. However, these distributions do not reflect the often significant differences in fit quality for a single shower. Furthermore, even if the wavefront shape were

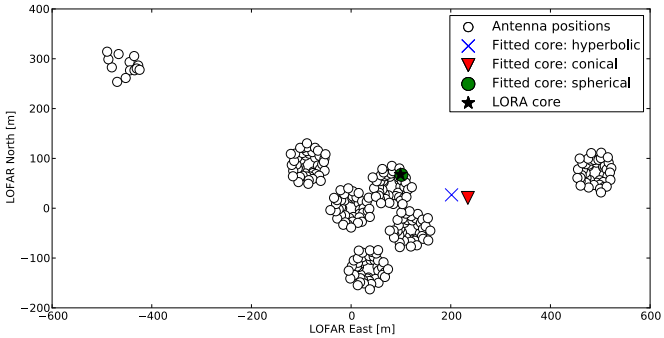


Figure 2.7: Fitted shower core positions for the shower in Fig. 2.5, for the different wavefront shapes. Note that the core position determined by the LORA particle detector array is not reliable for this particular air shower since it is located at the edge (or even outside) of the instrumented area.

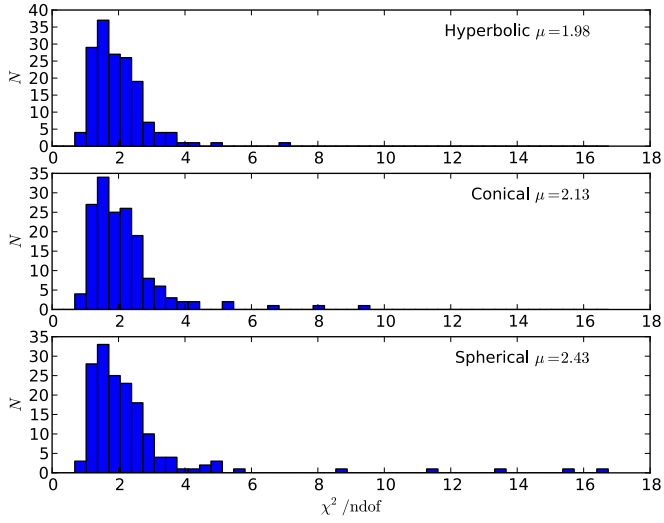


Figure 2.8: Fit quality for a hyperbolic (top), conical (middle) and spherical (bottom) wavefront shape.

always hyperbolic one would still expect to see shapes that appear conical or spherical for individual showers depending on the shower geometry and the part of the shower front that is sampled by the detector.

In order to check which wavefront shape is favored by the overall dataset we perform a likelihood ratio test. The test statistic for the conical case is:

$$D = -2 \frac{\ln(\text{likelihood hyperbolic})}{\ln(\text{likelihood conical})} \tag{2.11}$$

$$= \sum_{k}^{N} \chi^2_{\text{con}} - \chi^2_{\text{hyp}} \tag{2.12}$$

where the sum k is over all N showers. For an appropriate choice of parameters the hyperbolic function can turn into either a conical or a spherical function. Thus, the solution space of the spherical and conical fit functions are subsets of the solution space of the hyperbolic fit. Therefore (if the fit converged correctly) the hyperbolic fit will always have a lower χ^2/ndf value, even when the wavefront shape is intrinsically spherical or conical.

Under the null hypothesis that the wavefront shape is intrinsically conical (or spherical) the test statistic D should follow a $\chi^2(N)$ distribution. For large N, the $\chi^2(N)$ distribution approximates a Gaussian with mean N and standard deviation $\sqrt{2N} \ll D - N$. From the data we obtain the value $D = 6309$. The probability for this value to occur if the shape is conical is effectively zero, $p \ll 10^{-9}$, as the D-value is very far out of the distribution range.

There are two possible reasons for obtaining a higher value. Either the timing uncertainties are underestimated or the wavefront shape is generally not conical. Given the obtained reduced χ^2 values of the hyperbolic fit, averaging to 1.98, it is unlikely that the uncertainties are underestimated by more than a factor ~ 1.5. This is not enough by far to explain the measured value of the test statistic. Therefore we reject the null hypothesis that the wavefront shape is conical. Using the same procedure we also reject a spherical wavefront shape, with $D = 16927$ and correspondingly an even (much) lower p-value. Moreover, the lack of overall structure in the residuals of the hyperbolic fits (at our timing precision) argues against a more complicated wavefront shape. Therefore we conclude that the wavefront shape is hyperbolic. Furthermore, we do not see any evidence for a deviation from rotational

symmetry (around the shower axis). So this is either not present, or is not resolvable with the current timing resolution.

2.5.2 Direction reconstruction

Comparing the reconstruction of the shower axis direction for the best fitting hyperbolic wavefront to the planar wavefront solution, in the top panel of Fig. 2.9, we see that the latter deviates by up to $\sim 1\,^\circ$. Therefore, using a non-planar wavefront shape leads to an improvement in reconstruction precision of the air shower direction.

As can be seen from the bottom panel in Fig. 2.9 however, the exact shape of the non-planar wavefront is less important. The difference between the reconstructed direction using a conical or hyperbolic wavefront shape is typically less than $0.1\,^\circ$. Since a conical fit contains one less free parameter this may be more practical in reconstruction software. However, a planar fit does not depend on the position of the shower axis or the exact shape of the wavefront and is thereby more robust and more suitable for standard reconstruction software when higher precision is not required.

2.5.3 Correlations with air-shower parameters

From Schröder et al. (2011) it is expected that the shape of the radio wavefront depends on air-shower parameters and the distance to the shower maximum in particular. Since, for a shower with the same $X_{\max}$ the distance to shower maximum increases with increasing zenith angle (θ), the shape of the radio wavefront is also expected to depend on the zenith angle. This can be seen in Fig. 2.2 where the radius of curvature of the inner part, its extent and the slope of the conical part are all expected to depend on the distance to the last emission point. This in turn would depend on $X_{\max}$.

Similar to Schröder et al. (2011), we can take e.g. the time lag of the radio wavefront at $r = 100\,\mathrm{m}$, with respect to the arrival time of the emission along the shower axis ($r = 0$). It is not possible to use the hyperbola parameter b (the slope of the asymptote) directly, as in some cases the asymptotic regime is (far) outside the data range. Fig. 2.10 shows the time lag at $r = 100\,\mathrm{m}$ as a function of zenith

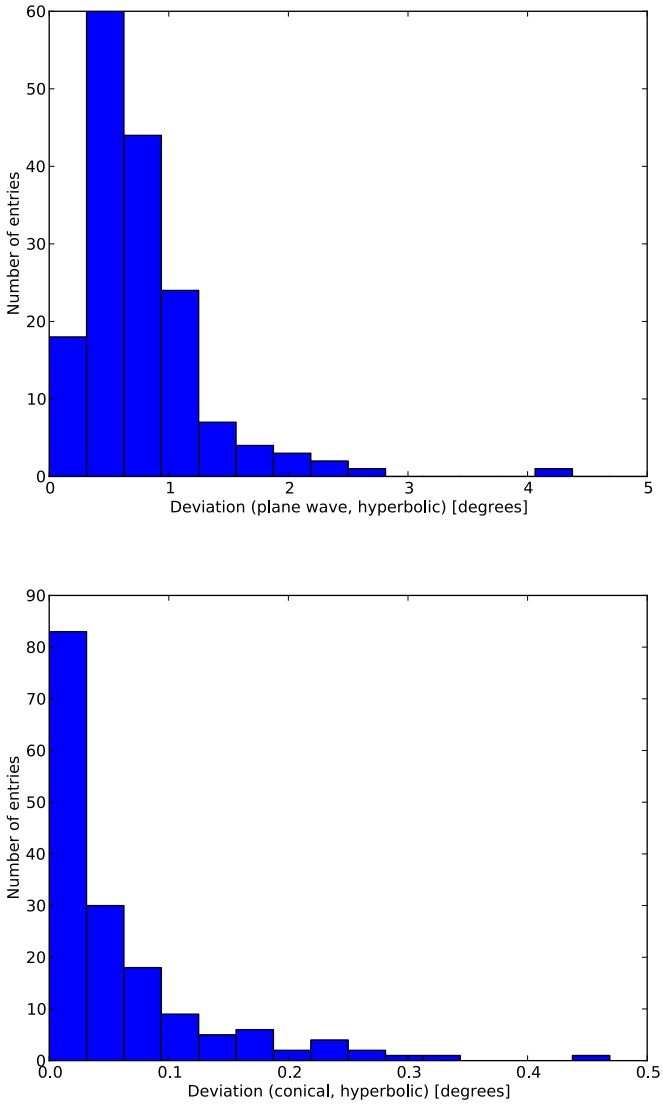


Figure 2.9: Angular difference between reconstructed shower axis direction for three wavefront shape assumptions. Assuming a planar wavefront shape typically introduces an error in the direction of up to $\sim 1^\circ$, when the shape is in fact hyperbolic (top plot). The differences in reconstructed direction between a conical and hyperbolic wavefront shape are approximately a factor of ten smaller (bottom plot).

angle. We find a weak correlation with a Pearson correlation coefficient of -0.32. The probability of obtaining this value for uncorrelated data is $4 \cdot 10^{-5}$.

To give an order of magnitude of the angular deviation between the measured wavefront and the shower plane, we can use t_{100} to get

$$\alpha = \frac{c\, t_{100}}{100\,\mathrm{m}}, \tag{2.13}$$

which is on average $0.011\,\mathrm{rad} = 0.63\,^\circ$. As the hyperbola becomes steeper further out, we could also use t_{250} instead (still inside the data range), which would give on average $0.94\,^\circ$. These numbers agree qualitatively with the average deviation angle from a plane of $0.83\,^\circ$ found by Schröder et al. (2011). The small angle of less than one degree explains why precise timing is required in order to measure the wavefront shapes.

In practice however, it appears to be difficult to use wavefront timing by itself to determine (the distance to) $X_{\max}$. This is due to the strong interdependency of the shower axis position and the exact shape of the wavefront. While the wavefront shape remains hyperbolic when moving the shower axis location around, the curvature near the axis as well as the slope further out change. Therefore it is best to combine timing information with other information available on the shower. This information may come from the particle detectors, or from the radio data in the form of the intensity pattern at ground level. It has already been shown that the radio intensity pattern itself is highly sensitive to $X_{\max}$ (Buitink et al., 2013). Combining this technique with timing information will improve the precision of these measurements.

2.6 Conclusions

We have shown that the wavefront of the radio emission in extensive air showers is measured to a high precision (better than 1 ns for each antenna) with the LOFAR radio telescope. The shape of the wavefront is best parametrized as a hyperboloid, curved near the shower axis and approximately conical further out. A hyperbolic shape fits significantly better than the previously proposed spherical and conical shapes.

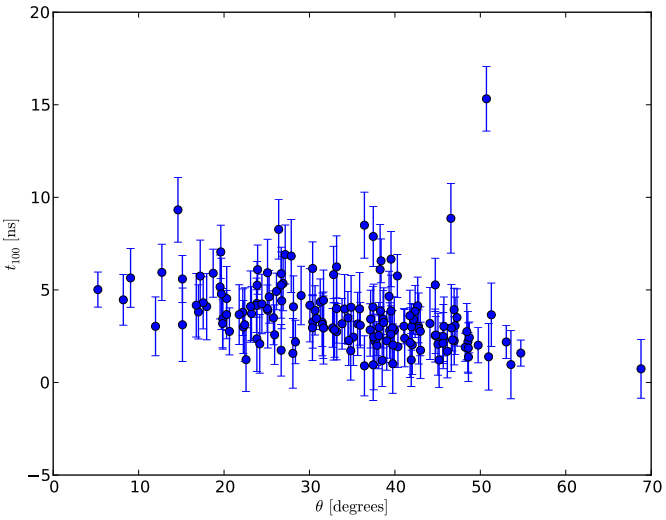


Figure 2.10: Time lag of the radio wavefront at $r = 100\,\mathrm{m}$, with respect to the arrival time of the emission along the shower axis ($r = 0$), as a function of zenith angle. Arrival times are obtained by evaluating the best fitting hyperbolic fit at $r = 100\,\mathrm{m}$. Uncertainties on t_{100} represent one standard deviation of the scatter around the best fitting hyperbolic fit over the full range of r, which are typically a factor ~ 1.5 larger than the timing uncertainty for individual antennas.

Reconstruction of the shower geometry using a hyperbolic wavefront yields a more precise determination of the the shower direction, and an independent measurement of the core position. Assuming the resulting reconstructed direction has no systematic bias, the angular resolution improves from $\sim 1^\circ$ (planar wavefront) to $\sim 0.1^\circ$ (hyperbolic). This assumption will be tested in a future simulation study. This improvement will be of particular importance for radio $X_{\max}$ measurements for highly inclined showers where small deviations in arrival angle correspond to large differences in the slanted atmospheric depth.

The high antenna density and high timing resolution of LOFAR offer a unique opportunity for a detailed comparison with full Monte Carlo air shower simulations, including the arrival time measurements presented here. Furthermore, efforts to integrate timing information within the $X_{\max}$ measurement technique from Buitink et al. (2013) are currently ongoing.

Acknowledgements

The authors thank Eric Cator for a useful discussion on statistical tests, and thank the anonymous referee for constructive comments. The LOFAR Key Science Project Cosmic Rays very much acknowledges the scientific and technical support from ASTRON.

Furthermore, we acknowledge financial support from the Netherlands Research School for Astronomy (NOVA), the Samenwerkingsverband Noord-Nederland (SNN) and the Foundation for Fundamental Research on Matter (FOM) as well as support from the Netherlands Organization for Scientific Research (NWO), VENI grant 639-041-130. We acknowledge funding from an Advanced Grant of the European Research Council under the European Union's Seventh Framework Program (FP/2007-2013) / ERC Grant Agreement n. 227610. Chiara Ferrari acknowledges financial support by the *"Agence Nationale de la Recherche"* through grant ANR-09-JCJC-0001-01.

LOFAR, the Low Frequency Array designed and constructed by ASTRON, has facilities in several countries, that are owned by various parties (each with their own funding sources), and that are collectively operated by the International LOFAR Telescope (ILT) foundation under a joint scientific policy.

Chapter 3

Timing calibration and spectral cleaning of LOFAR time series data

A. Corstanje et al. (LOFAR Key science project Cosmic Rays)

Astronomy and Astrophysics 590, A41 (2016)

Abstract

We describe a method for spectral cleaning and timing calibration of short time series data of the voltage in individual radio interferometer receivers. It makes use of phase differences in fast Fourier transform (FFT) spectra across antenna pairs. For strong, localized terrestrial sources these are stable over time, while being approximately uniform-random for a sum over many sources or for noise. Using only milliseconds-long datasets, the method finds the strongest interfering transmitters, a first-order solution for relative timing calibrations, and faulty data channels.

No knowledge of gain response or quiescent noise levels of the receivers is required. With relatively small data volumes, this approach is suitable for use in an online system monitoring setup for interferometric arrays.

We have applied the method to our cosmic-ray data collection, a collection of measurements of short pulses from extensive air showers, recorded by the LOFAR radio telescope. Per air shower, we have collected 2 ms of raw time series data for each receiver. The spectral cleaning has a calculated optimal sensitivity corresponding to a power signal-to-noise ratio of 0.08 (or -11 dB) in a spectral window of 25 kHz, for 2 ms of data in 48 antennas. This is well sufficient for our application. Timing calibration across individual antenna pairs has been performed at 0.4 ns precision; for calibration of signal clocks across stations of 48 antennas the precision is 0.1 ns. Monitoring differences in timing calibration per antenna pair over the course of the period 2011 to 2015 shows a precision of 0.08 ns, which is useful for monitoring and correcting drifts in signal path synchronizations.

A cross-check method for timing calibration is presented, using a pulse transmitter carried by a drone flying over the array. Timing precision is similar, 0.3 ns, but is limited by transmitter position measurements, while requiring dedicated flights.

3.1 Introduction

An interferometric radio telescope relies on an accurate timing calibration of the signals of all its constituent receivers, in order to be able to combine signals with a time or phase shift corresponding to the direction of a given source in the sky. Furthermore, spurious narrowband transmitter signals, which are present even in relatively radio-quiet regions, will also show up in the processed signals. These signals have to be identified and removed, preferably early in the analysis process.

This paper is organized as follows: below we briefly review some methods that are used for detection and removal of radio-frequency interference (RFI), as well as

methods for timing and phase calibration. After this, we introduce our methods. Section 3.2 describes the methods in detail, and in Section 3.3, their application to data taken with the LOFAR radio telescope is discussed.

Most of the radio-frequency interference (RFI) present at radio telescope sites consists of either narrowband signals from radio transmitters, or short pulses in the time domain (Offringa et al., 2013). For both cases, there are several methods being used to identify interference, either before or after signal correlation. Before correlation in the interferometer, these algorithms typically involve detecting threshold crossings of amplitudes in the time or frequency domain, where the threshold is adapted based on signal properties (Offringa et al., 2010). For instance, a threshold can be calculated by using a median filter, such as that used in the Auger Engineering Radio Array (AERA) for radio detection of cosmic rays (Schmidt et al., 2011). It replaces a sample in time or frequency domain by the median of a number of its neighbours in order to set a threshold. More elaborate techniques, also exploiting correlations of multiple samples crossing the threshold, are found e.g. in Offringa et al. (2010). After correlation, adaptive thresholds can also be used on correlated visibility amplitudes instead of data streams from single receivers.

Another approach that has been considered for use in the AERA experiment is described in Szadkowski et al. (2013). It uses linear prediction, implemented as a finite impulse response (time domain) filter in FPGAs. This approach operates online on single receivers and adapts to changes in the interference environment.

Another method is *fringe fitting*, which makes use of the fact that most RFI sources are at a fixed position, and therefore produce sinusoidal fringes in visibility data of a fringe-stopping interferometer (Athreya, 2009). These sinusoids are then fitted and removed. This latter method is somewhat similar to the method we present below, which operates on short time series.

Timing and phase calibration in interferometric radio telescopes is typically done based on the principle of self-calibration (Pearson and Readhead, 1984; Taylor et al., 1999), where redundant information in the interferometric data is used; for instance, there are $N_{\mathrm{ant}}(N_{\mathrm{ant}}-1)/2$ baselines giving correlated signals, while having only N_{ant} antennas to calibrate. For this method, suitable calibrator sources for which the structure is known, e.g. point sources, are used as a model for optimizing the calibration. The calibration solution can be obtained as a function of frequency,

providing a phase calibration for every frequency in the spectrum. The phase calibration at a given frequency equals a timing calibration at the same frequency, taken modulo the wave period.

Moreover, there are methods that also take into account directional dependencies of the calibration. As antennas have a complex gain that has directional dependence, the calibration in general depends on this as well, especially considering differences in gain between antennas. One of these methods, which is used at LOFAR, is SAGECal (Kazemi et al., 2011). A review of similar calibration methods is given in e.g. Wijnholds et al. (2010).

Alternative approaches involve calibrating on a fixed custom transmitter, as is done by the LOPES cosmic-ray detection experiment (Schröder et al., 2010), which yields a timing calibration per antenna for a single, or a few frequencies. In our approach, as described below, we use the spectral cleaning method to identify a suitable public transmitter, and also make use of the position of the most useful transmitter in order to obtain a calibration solution. This is sufficient for a precise (subnanosecond) timing analysis of cosmic ray pulses (Corstanje et al., 2015). It can also serve as a starting point and cross-check for dedicated phase calibrations as used in radio astronomy.

Instead of a fixed transmitter, satellites or drones flying overhead can be used, which makes amplitude calibration possible as well. This is similar to the amplitude calibration from a fixed transmitter as has been performed at LOFAR (Nelles et al., 2015a).

Calibration on pulses from the far field, e.g. emitted by airplanes passing overhead, has also been considered (Aab et al., 2016b). However, this relies on randomly occurring pulses that need to be detected in real-time in order to record them.

Here, we describe a method of spectral cleaning of time series data that we use to remove narrowband radio-frequency (RF) transmitter signals from our data. At the same time it allows a calibration of clock differences to be obtained across the array. The method applies only to narrowband signals, which are present continuously for about 0.2 to 2 ms, where shorter signals need to be stronger to be detected. Signals with somewhat larger bandwidth are treated as a set of narrowband signals. Broadband pulses are not removed.

Using the phase component of the Fourier transform of each channel, we take

into account that strong, localized transmitters produce approximately constant phase differences across the array. Astronomical signals are typically broadband, and arrive at the antennas as a sum over many sources on the sky, and therefore produce random phase differences over time. This difference allows for an accurate identification (and removal) of disturbing signals. Using the identified constant phases of a public radio transmitter signal, we can also calibrate signal timing offsets in each antenna pair. If the geometric delay from the signal path lengths of the radio signal to each antenna is known, this leads to a known difference in phase of the (continuous-wave) signal as it is measured at each antenna. Comparing the actually measured phases with the expected phases gives a calibration correction. It has been suggested as a promising improvement in Offringa et al. (2010) to add the use of phase information to existing amplitude-based RFI cleaning methods. The method presented here uses only the phase component.

We apply this method to data taken with the Low Frequency Array (LOFAR) (van Haarlem et al., 2013) radio telescope. The antennas of LOFAR are distributed over northern Europe, with the densest concentration in the north of the Netherlands. The antennas are organized into stations, each consisting of 96 low-band antennas (LBA, 10 - 90 MHz), and 48 high-band antennas (HBA, 110 - 240 MHz). Within the core region of about $6\,\mathrm{km}^2$, 24 of these stations have been distributed.

For the cosmic-ray data collection, we record radio emission from extensive air showers, which reaches the ground as a short pulse, on the order of 10 to 100 ns long (Schellart et al., 2013). We use the Transient Buffer Boards installed in the data channel of every LOFAR antenna to record these pulses and other fast radio transients. Each recording is 2 to 5 ms long and consists of the raw voltage time series from every data channel. The buffer is capable of storing signals up to 5 seconds in length.

These datasets need spectral cleaning in order to measure the pulses accurately. The relative timings of the pulses contain information about the air shower process. For instance, by measuring pulse arrival times, we have evaluated the shape of the radio wavefront as it arrives at the antenna array (Corstanje et al., 2015).

As our datasets are much shorter than typical astronomical observations (a few milliseconds, instead of hours), and are stored as unprocessed voltage time series per receiver, a dedicated spectral cleaning method is required. Still, our method

can be easily adapted for other purposes and instruments as long as raw time series are available.

We tested our timing calibration using a pulse transmitter carried by an octocopter drone flying above the array. The precision of the pulse arrival time measurements is similar to the phase measurements.

3.2 Method

In this section we explain in detail the method and performance of our RFI identification algorithm, and show how the phases of the strong transmitters identified in this way can be used for timing calibration.

3.2.1 Radio frequency interference identification

In order to identify frequencies that are contaminated by human-made interference, a typical approach is to search for strong signals above the noise level in an amplitude or power spectrum. However, this requires knowledge of the noise power spectra in the absence of RFI transmitters, or an adaptive or iterative technique to estimate them, as mentioned in the Introduction.

Therefore, we use the relative phases between pairs of antennas. At the frequency used by a transmitter, the phase difference across an antenna pair is stable over time because the signals are typically transmitted from a fixed location. In contrast, at frequencies where no terrestrial transmitter is present, we measure emission from the Milky Way as well as electronic noise. The Galactic emission is a sum of many sources, assuming the antennas are omnidirectional or have a substantial field of view. Therefore, the detected phases can be treated as random on millisecond integration timescales.

In situations where one localized source in the sky fully dominates the signal, for example during strong solar bursts, this assumption is not valid. However, this only happens for a small fraction of the time.

We take phase measurements from a fast Fourier transform (FFT) of consecutive data blocks for every antenna. One antenna is taken as reference; for every frequency channel, its phase is subtracted to measure only relative phases.

It is also possible to consider the phase differences across all antenna pairs (baselines), instead of selecting a single antenna as reference. This is more sensitive (see Sect. 3.2.2), but also requires more computation time, and hence can be omitted if the single-reference approach meets the requirements for spectral cleaning.

For every frequency channel we calculate the average and variance of the phase over all data blocks. The phase average across antenna indices j and k for frequency ω is defined as follows (denoting relative phases as Φ and the data block number as superscript l),

$$\Phi^l_{j,k}(\omega) = \phi^l_j(\omega) - \phi^l_k(\omega), \tag{3.1}$$

$$\bar{\Phi}_{j,k}(\omega) = \arg\left(\sum_{l=0}^{N_{\mathrm{blk}}-1} \exp(i\,\Phi^l_{j,k}(\omega))\right), \tag{3.2}$$

and the phase variance as

$$s_{j,k}(\omega) = 1 - \frac{1}{N_{\mathrm{blk}}}\left|\sum_{l=0}^{N_{\mathrm{blk}}-1} \exp(i\,\Phi^l(\omega))\right|, \tag{3.3}$$

where $\Phi^l(\omega)$ is the relative phase measured in data block l at frequency ω and N_{blk} is the number of data blocks. The phase variance $s_{j,k}(\omega)$ is close to unity for completely random phases and is zero for completely aligned phases.

If the phases follow a narrow, peaked distribution around the average with variance σ^2, then the phase variance is $s_{j,k} = \frac{\sigma^2}{2} + \mathcal{O}(\sigma^4)$. Hence, this quantity is then indeed proportional to the variance. For wider distributions, the 2π-periodicity of phases becomes important, and the phase variance has a maximum value of unity for a uniform distribution, in the large-N limit.

For random phases, $N_{\mathrm{blk}}(1 - s_{j,k})$ describes the traveled distance in a two-dimensional random walk as the right-hand part of Eq. 3.3 represents the length of the sum-vector of N_{blk} unit vectors, each of which has a random direction in the complex plane.

For large N_{blk}, this distance follows a Rayleigh distribution with scale parameter $s = \sqrt{N_{\mathrm{blk}}/2}$ (Rayleigh, 1905) and has an expected value of $\alpha\sqrt{N_{\mathrm{blk}}}$, with $\alpha = \frac{1}{2}\sqrt{\pi} \approx 0.89$. It has a standard deviation of $\beta\sqrt{N_{\mathrm{blk}}}$, with $\beta = \sqrt{1 - \frac{\pi}{4}} \approx 0.46$. In practice, the large-N approximation is already accurate for $N_{\mathrm{blk}} \gtrsim 10$.

Therefore, we have

$$s_{j,k}(\omega) \approx 1 - \frac{\alpha}{\sqrt{N_{\mathrm{blk}}}}. \tag{3.4}$$

It is clear that for a coherent, narrowband signal seen at all antennas, the variance should be rather small.

In order to determine a threshold for significantly detecting a transmitter, we take the average of the phase variances over all antennas or all baselines. This leaves one averaged phase-variance spectrum, i.e. one phase variance per frequency channel. To take the average is a simple, generic choice; when partial detections are expected, e.g. in a more sparse array or for very nearby RFI, it is possible to consider the full distribution of the phase variance over the antennas and test for deviations of random behavior. This is, however, not pursued here.

We sort the values of the phase variance over all frequencies, and estimate its standard deviation by taking the 95th percentile value minus the median, which is about $1.65\,\sigma$ for Gaussian noise. This has the advantage of considering only the upper half of the sample which is assumed to follow the random-walk characteristics. It selects out all transmitter signals as they only lower the variance below the median. This naturally assumes that less than half of the frequency channels contain an interfering transmitter signal, which is reasonable for astronomical observations in general. Should this not be the case for the particular site, a higher percentile value can be taken instead of the median. Alternatively, it is possible to follow the random-walk statistics directly for mean and standard deviation, to compare them with the data afterwards. The threshold is set to the median value minus a multiple of the standard deviation, which is tunable, for example, to trade a lower false-positive probability for a lower sensitivity.

For the run-time complexity, we note that the algorithm requires $N_{\mathrm{blk}}\,N_{\mathrm{ant}}$ FFTs of a fixed length, set by the desired spectral resolution. Moreover, when treating all baselines, it requires $\mathcal{O}(N_{\mathrm{ant}}^2)$ phase spectrum comparisons. When instead using a single antenna as reference (or a fixed number of them), only $\mathcal{O}(N_{\mathrm{ant}})$ comparisons are done, and the FFTs always dominate.

3.2.2 Sensitivity of RFI detection

The sensitivity of RFI detection can be analyzed by noting the correspondence of the detection of an RFI transmitter to the detection of bias, i.e. a preference towards a certain direction, in a set of identically distributed random walks. The full analysis is deferred to the Appendix.

We start by assuming that a signal is present in the noisy time series of each receiver, with power signal-to-noise ratio S^2 defined by the absolute-square of the FFT in one frequency channel, for the signal and the noise, respectively.

With this definition, the sensitivity becomes (asymptotic approximation, see Appendix)

$$S^2 > 3.8\,\sqrt{k/6}\,N_{\rm blk}^{-1/2}\,N_{\rm ant}^{-1/2}, \tag{3.5}$$

aimed at a six-sigma ($k = 6$) detection threshold as used in the LOFAR cosmic-ray analysis (Schellart et al., 2013). Typical numbers for this analysis are $N_{\rm blk} = 50$ and $N_{\rm ant} = 48$, leading to a threshold of $S^2 = 0.08$, or -11 dB, which is easily sufficient for the purpose of analyzing pulses from air showers. With a specific bandwidth fraction in mind, e.g. for wideband radio signals, this sensitivity can be used to give an upper bound to residual RFI levels.

To put this result into perspective, we consider a straightforward, simplified method that detects excess power in a power spectrum, averaged over all antennas and data blocks. The noise power in a given FFT channel has an exponential distribution (Papoulis and Pillai, 2002) with mean and standard deviation equal to the mean noise power per channel. The signal power is uncorrelated to the noise, and hence the total power is the sum of signal and noise power. Summing spectra of $N_{\rm ant}$ antennas each having $N_{\rm blk}$ blocks then yields a threshold

$$S^2 = 6\,(k/6)\,N_{\rm blk}^{-1/2}\,N_{\rm ant}^{-1/2}, \tag{3.6}$$

plus the uncertainty in determining the average quiescent noise spectrum, which is not flat in general. The asymptotic behavior is therefore the same as in Eq. 3.5. It is assumed that estimating a single, flat noise level as for the phase variance (Eq. 3.4) has a lower uncertainty than estimating a spectrum curve.

We note that, since both methods use averaging over many blocks, or many phase variance values, by the Central Limit Theorem these averages can be regarded as estimating the mean of a Gaussian distribution. Hence, in both cases the $k-$sigma thresholds refer to exceeding probabilities, and corresponding false alarm rates, of a Gaussian distribution.

Our method based on phases has a somewhat favorable detection threshold, the difference with respect to Eq. 3.6 being at least 2.0 dB. Moreover, it does not require an estimate of the noise spectrum in the absence of transmitters. This has

made it easier for us to implement in practice where background levels are variable. However, this does not imply that this comparison holds when looking at more elaborate, amplitude-based cleaning methods.

We note that when, instead of power excess, the amplitude excess would be used in an absolute spectrum rather than absolute-squared, the asymptotics of S are the same, i.e. given by Eq. 3.6. In the weak-field and large-N limit, the difference is at most 7 % in the constant factor in favor of the absolute spectrum.

3.2.3 Timing calibration

Observations using an interferometric telescope require precise timing and phase calibration of each receiver in order to have precise pointings and to perform imaging with optimal signal quality. The timing precision should be about an order of magnitude below the sampling period.

For timing calibration we use one or multiple narrowband transmitters as a beacon, producing fixed relative phases between antennas at the transmitting frequency. This is similar to the procedure followed in Schröder et al. (2010); we extend this by a more precise phase measurement, and by using the geometric delays from the transmitter location to find the calibration delays per antenna pair.

We measure relative phases per antenna pair in the same way as in Sect. 3.2.1, taking Fourier transforms of consecutive data blocks for all antennas and averaging phases using Eq. 3.2. This also allows frequencies suitable for timing calibration to be identified from the values of the phase variance, Eq. 3.3, where lower values are better.

The geometric delays from the transmitter to each antenna are needed to determine the calibration delays between antennas. Therefore, a transmitter at a known location must be used with a frequency above about 30 MHz. At lower frequencies, i.e. the HF-band (High Frequency), signals may reflect off the ionosphere and propagation characteristics may vary from time to time; see e.g. Gilliland et al. (1938).

The signals propagate with the light speed in air, which is c/n. The index of refraction n is on average 1.00031; we note that variations of $\pm\, 4\, 10^{-5}$ are possible with temperature and humidity (Grabner and Kvicera, 2011). Omitting the refractive index would therefore introduce a timing mismatch of 1.0 ns between two

antennas separated by 1 km, along the line of sight to the transmitter. This value is therefore significant at intermediate and longer baselines.

The phase difference across a given antenna pair, after accounting for geometric delays from the transmitter, corresponds to a time difference (calibration mismatch)

$$\Delta t = \frac{\Delta\phi}{2\pi f} \pmod{\frac{1}{f}}. \tag{3.7}$$

Thus, the calibration solution obtained by using one transmitter is only determined up to a multiple of the signal period $1/f$. This can be improved by combining results from multiple transmitters. However, in order to obtain the correct solution it is then required that the different transmitters have larger differences in period than the phase or timing noise. Moreover, in general the correct calibration phase depends on frequency, i.e. the optimal phase calibration may have deviations from the group delay, as a function of frequency. When using a custom beacon for calibration measurements as described here, it is preferred therefore to choose frequencies that are far apart.

Once antenna timings have been calibrated, the relative phases can be monitored over time without reference to the transmitter location and geometric delays.

3.3 Application to LOFAR data

In this section, we describe how the RFI identification and timing calibration methods are used for the analysis of air shower datasets with LOFAR.

3.3.1 Identification of radio-frequency interference

The LOFAR radio telescope, located in the north of the Netherlands, is in a relatively radio-quiet region. Nevertheless, in all observations there are several signals present that come from narrowband transmitters. Therefore, spectral cleaning methods are required to remove them from astronomical observation data.

We have used the core stations of LOFAR. For all but a few very bright air showers, our data contain antenna baselines up to about 1 km, and most of the antennas with signal are in the central ring that is 320 m in diameter. An example power spectrum is shown in Fig. 3.1. For demonstration purposes, the dataset of this example has particularly bad RFI as there are several flagged frequencies in

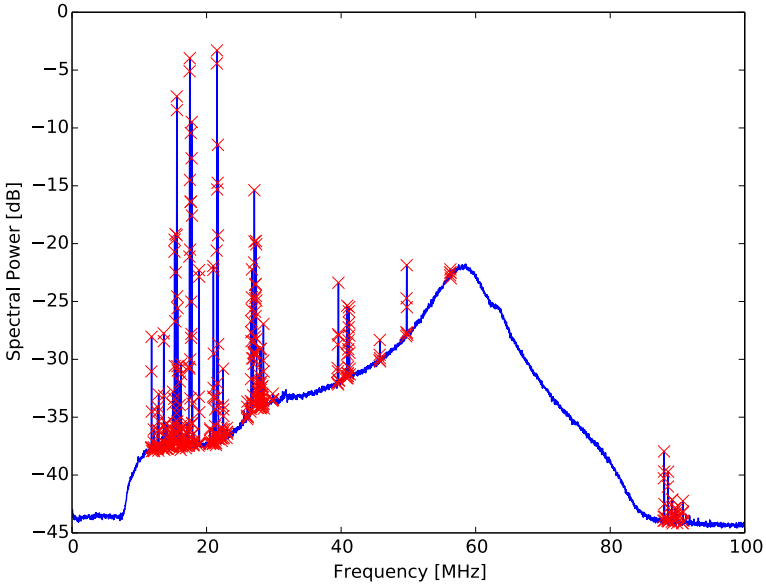


Figure 3.1: Example power spectrum from 2 ms of LOFAR data, averaged over 48 LBA antenna dipoles. Crosses indicate channels with detected transmitters.

the 30 to 80 MHz-band. However, this case is still not too extreme, and spectral cleaning is indeed necessary in similar instances. The power spectrum is averaged over the 48 antennas in one LOFAR station, and averaged over 2 ms, which is the length of a typical cosmic-ray dataset. We treat each of the two instrumental polarizations separately as RFI signals may be detectable in only one of the two polarizations. It is a spectrum of the LBA antennas ranging from 10 to 90 MHz. In what follows we focus only on the low-band spectra as these are best used for air shower measurements; the methods work identically for the high-band antenna data. For the detection of the transmitter frequencies, we use FFTs with a block size of 8000 samples, which amounts to a spectral resolution of 25 kHz. There are then 50 blocks in a time series of 2 ms, which are used to calculate the phase variance over the entire 2 ms of data as in Eq. 3.3. The result is shown in Fig. 3.2. The

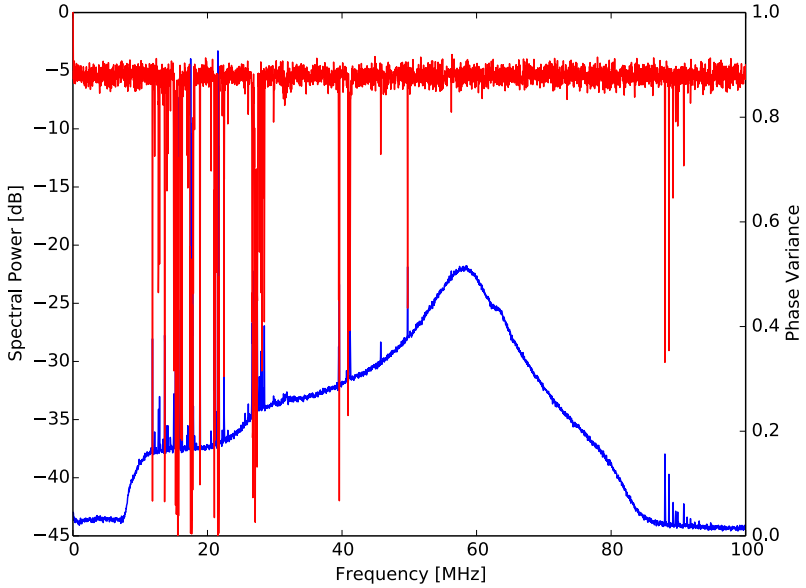


Figure 3.2: Example power spectrum from 2 ms of LOFAR time series data (lower curve). The phase variance is shown in the upper data points (red). It consistently becomes lower whenever a narrowband transmitter is seen in the power spectrum.

phase variance, taken as the median value of the 48 antennas, is shown as the upper signal. It has random noise due to the finite number of data blocks; at frequencies where a narrowband transmitter is present in the power spectrum (lower curve), the variance is significantly lower. The random noise has a median value of 0.879, consistent with the expected value from Eq. 3.4 of 0.875 for 50 data blocks. This is a basic test of our randomness assumption for the phases.

The phase variance threshold is then set to a value of nearly six sigma. The standard deviation is estimated by the 95th percentile value minus the median, which is about $1.65\,\sigma$ for Gaussian noise. Every frequency channel with lower phase variance is flagged.

When frequency resolution (set by the chosen FFT block size) is high enough

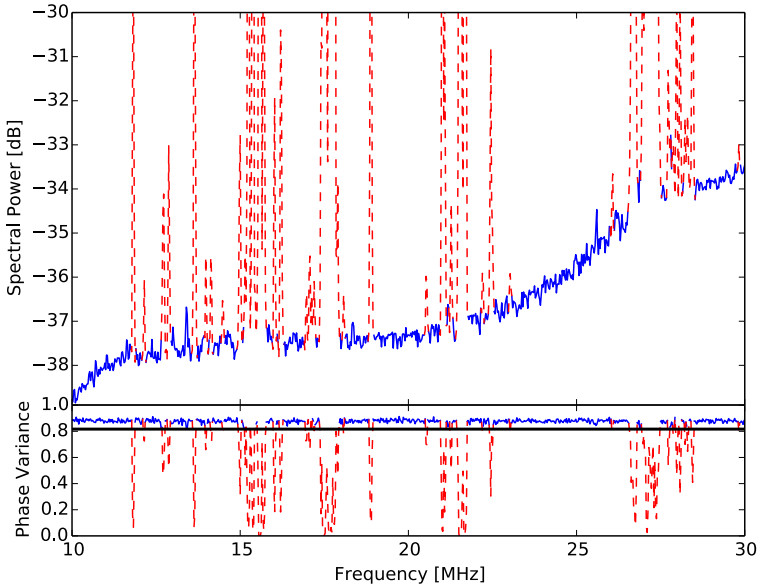


Figure 3.3: Close-up of the power spectrum in a frequency range with several RFI sources. Flagged frequencies are shown as red dashed lines. The lower panel shows the phase variance, where the black horizontal line indicates the threshold for flagging. Although the RFI-quiet noise level would follow a smooth curve, fitting the curve and RFI flagging using the excess power are interdependent.

to resolve the transmitters' frequency responses, it can also be necessary to flag a number of adjacent frequency channels as the edges of resolved transmitter spectra may not meet the threshold criterium for flagging. This is especially important when a large block size is taken, e.g. to comply with FFT resolution used in later analysis. The number of adjacent channels to flag is currently set as a manually tunable parameter, scaling with frequency resolution.

In Fig. 3.3, a close-up of the power spectrum and the phase variance are shown.

3.3.2 Timing calibration: results for the LOFAR core

For calibration of short time series, i.e. 2 to 5 ms in length, we use one or multiple narrowband transmitters as a beacon, producing fixed relative phases between antennas at the transmitting frequency. The signals at the high end of the spectrum (> 87 MHz) are from public radio transmitters which are always present. They are clearly detectable even though they are outside the passband of the filters, which ends at 80 MHz. Moreover, the phase variance we measure in the spectral cleaning algorithm, and the corresponding timing precision, is found to be best for these frequencies. Therefore, we work with the high-frequency transmitters, especially the strongest one at 88.0 MHz. The radio signals at frequencies 88.0, 88.6, 90.8, and 94.8 MHz are transmitted from a 300-meter radio tower located in Smilde[1] at 31.8 km from the LOFAR core. For 88.0 MHz, the signal period is 11.3 ns, which is still much larger than the desired (and attainable) subnanosecond calibration precision.

The timing calibration signal follows from the relative phases after accounting for the geometric delays between transmitter and antennas, according to Eq. 3.7. The relative phases are once again obtained from the FFT of 50 consecutive data blocks, taking average phases as from Eq. 3.2. As was done for the RFI detection method, we treat the two polarizations of the LOFAR LBA antennas separately. We thereby make use of the identical design and orientation of the LOFAR antennas. If antenna orientations and/or the design of their polarizations are different, this could lead to larger timing errors in this procedure, when using transmitters with polarized signals. Monitoring of a cross-calibration over time would still be accurate (see Sect. 3.3.3 below).

The geometric delays are calculated using the International Terrestrial Reference Frame (ITRF) coordinates (Altamimi et al., 2002) of each antenna, and the GPS (WGS-84) (Defense Mapping Agency, 1987) ellipsoid coordinates of the Smilde tower converted to ITRF. This is a cartesian coordinate system, allowing for an easy calculation of straight-line distance between two points. For the effective height of the emission we consider half the height of the tower; the uncertainty in relative timings per 100 m of height is less than 0.05 ns across LOFAR core stations and below 0.005 ns within one station, and therefore negligible for our purposes.

[1] GPS coordinates: 6.403565 ° east, 52.902671 ° north.

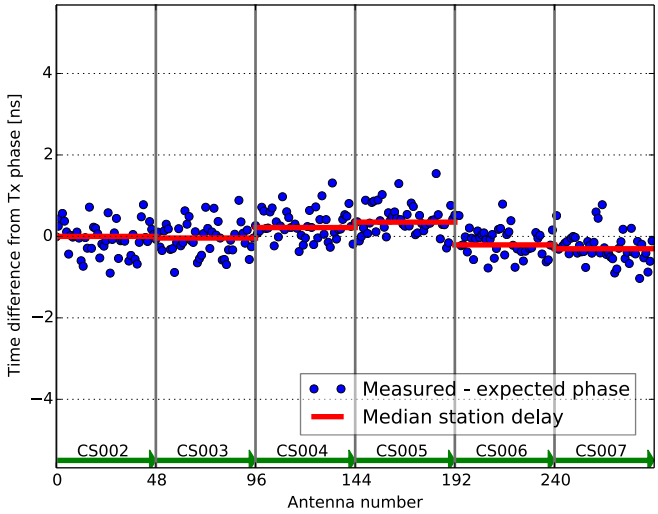


Figure 3.4: Difference between measured and expected phases per antenna, converted to time in nanoseconds. Red solid bars represent the median time delay per LOFAR station (48 antennas); the stations are separated by the vertical grid lines. The range of the y-axis corresponds to the signal period at 88 MHz.

As a starting point we take an existing LOFAR timing calibration per antenna, which is performed using astronomical phase-calibration a few times a year (van Haarlem et al., 2013). We compare measured relative phases with those from the straight-line propagation, in the LOFAR core area, consisting of a circular area 320 m in diameter, plus some additional stations up to about 1 km away. There are many more stations, but our air shower measurements are limited to this area.

The phases correspond to a timing correction per antenna as shown in Fig. 3.4. The values depicted in this plot consist of both calibration errors and possible systematic effects from our measurement. The latter may include differences in filter characteristics at 88.0 MHz, i.e. the delays obtained from phases at this frequency may deviate from the full group delay. Wave propagation effects may vary slightly over antennas, e.g. owing to the presence of other LOFAR antenna(s) along the line

of sight to the transmitter.

We can assume that any calibration mismatch with respect to the earlier LOFAR calibration is independent of these systematic effects. Dedicated calibration observations use astronomical sources instead of a terrestrial transmitter and span the entire frequency band. It is important to note, therefore, that the timing correction signal we find here provides an upper limit on both calibration errors and systematic effects.

The standard deviation of the timing correction signal is 0.44 ns. Per station, the standard deviation varies from 0.36 to 0.40 ns.

Our measurements and data collection started in June 2011, which was within the commissioning period of LOFAR; the 'cycle 0' observations started in December 2012. This means that some technical timing issues (which were resolved later) were still present during the early observations. Using this method, these problems were detected in the same datasets that contain our cosmic-ray measurements and have since been corrected. Hence, our older data can also be fully used.

As LOFAR is divided into separate stations, timing calibration across stations is also required. Especially before October 2012, only the six innermost stations had a common clock, but all other core stations had their own clock synchronized by GPS. This caused clock drifting across stations on the order of 10 ns, which is much longer than interferometric accuracy requirements.

Therefore, we calculated the inter-station clock offsets by taking the median of the time delays per antenna in each station. Using the median instead of the mean is more robust against calibration errors or malfunctioning of a small fraction of antennas. On the other hand, the median has a higher uncertainty for estimating the mean than taking the average. Still, taking the median is useful when batch-processing thousands of datasets.

When inter-station clock offsets vary by more than the signal period of 11.3 ns, they are still known accurately up to a multiple of this period. For the cosmic-ray pulse timing measurements as performed in Corstanje et al. (2015), the actual solution can be identified by using fits of the incoming direction of the radio pulse of the air shower. These fits are done on single-station level and hence are not influenced by the inter-station offsets.

The standard error of the median over one station amounts to 0.08 ns, and is

a factor of $\sqrt{\pi/2} \approx 1.25$ higher than the standard error of the mean. Therefore, the inter-station clock offsets can be determined to about 0.1 ns precision, assuming that the systematic effects average out over the antennas of each station.

3.3.2.1 Multiple transmitters for calibration

The calibration solution obtained from using one transmitter is only given up to a multiple of the signal period. This can be improved by combining results from multiple transmitters. However, to obtain the correct solution, it is required that the different transmitters have larger differences in period compared to the phase or timing noise. For the LOFAR environment, the difference in period between 88.0 and 90.8 MHz is only 0.35 ns, which is not always above the timing noise. The transmitter at 94.8 MHz is not as reliable because its signal is rather weak. Moreover, in general the correct calibration phase depends on frequency, i.e. the optimal phase calibration may have deviations from the group delay as a function of frequency. This leads to an additional source of uncertainty when combining multiple frequencies.

When instead using a custom beacon for calibration measurements, in the way we described here for the public radio signals, it would be better to choose frequencies that are farther apart. Differences in phase delay versus group delay may also show up in this case. Another option is to use a beacon sending short pulses or bursts, as we show in the next section. These pulses are not affected by periodicity.

3.3.2.2 Pulse arrival times from an octocopter drone

As a cross-check, we performed a pulse arrival time measurement in the LOFAR inner core region using a pulse transmitter mounted below an octocopter drone. The octocopter flies along a pre-programmed flight path using GPS coordinates. We set it to fly above the central antenna of the six innermost stations of LOFAR. A pulse of approximately 250 V is then transmitted every 8 μs from a height of about 50 m. The incoming signal is recorded using the Transient Buffer Boards. The individual pulses are timed by interpolating the time series using up-sampling, and taking the time of the first positive maximum after the signal exceeds a given threshold, set as a fraction of its amplitude. This method was found suitable for timing relatively long pulses with a broad maximum. The rise time of the pulses was

on the order of 50 ns, corresponding to about 3 periods at the resonance frequency of the LBA antennas, near 58 MHz. The pulses showed a strong signal-to-noise ratio in all the antennas we used, hence it was possible to identify the correct maximum for timing.

Geometric delays follow from a straight-line path from the pulse transmitter to each antenna; the calibration signal for each antenna pair is the remaining time delay after accounting for the geometric delays.

The actual position of the octocopter can vary depending on wind and flight control uncertainties. In order to determine the transmitter location more precisely at the time of the measurement, an optimization procedure was performed. The calibration signals have been minimized with respect to a given calibration of LOFAR, which for the majority of the antennas has an uncertainty of at most $\sigma = 0.4$ ns as shown in Sect. 3.3.2. The position shifts by the optimization procedure were found to be about 1 to 1.5 m, which is significant for timing purposes when calibrating from scratch. The fit uncertainty then depends nontrivially on the calibration delays themselves.

Comparing pulse arrival times at each antenna with the expected geometric delay of the signal path from transmitter to receiver, we obtain the calibration signal shown in Fig. 3.5. The calibration signal is an average over 10 pulses. The standard deviation of the timing calibration signal amounts to 0.26 ns. This is comparable to the result of 0.44 ns obtained using continuous-wave radio transmitters. Nevertheless, there is still some structure visible in the arrival times for one of the stations, labeled CS003. This may point to a non-optimal fit for the transmitter position.

3.3.3 System monitoring

We have monitored the relative delays between antennas over the course of nearly four years, comparing the results of the given procedure for all datasets in our collection. With at least one calibration at a given date, for which we also know the relative phases, the time variations can be monitored without reference to the transmitter location, wave propagation etc. Only the measured relative phases need to be compared.

A typical time variation plot is given in Fig. 3.6. Timing corrections have been binned, using one bin per day. The given uncertainties are the standard deviations

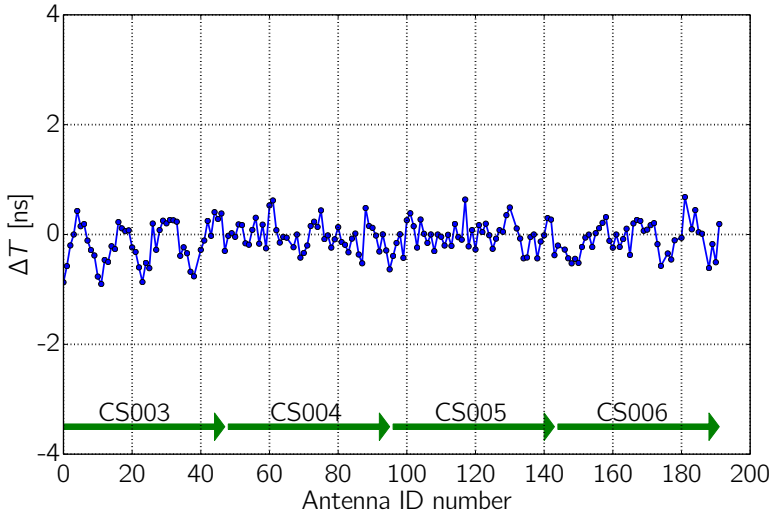


Figure 3.5: Arrival times of pulses from the octocopter drone. The points show the difference between measured and expected pulse arrival time per antenna for four of the innermost LOFAR stations indicated by the labeled arrows at the bottom.

over one day. The median value of this uncertainty is 0.08 ns, taken only from those days where at least five measurements were taken. This median uncertainty is also assigned to data points from days with less than five measurements. The relative timing between these two antennas is mostly stable over time at the 0.5 ns level, except for the first month of measurements which was within the commissioning time of LOFAR. After this, there were only three days when no stable solution for the timing was found; they are shown as large uncertainties in Figs. 3.6 and 3.7.

Figure 3.7 shows a close-up of the same plot. It shows slow clock drifting, and demonstrates that signal path synchronization at the level of 0.1 ns can indeed be followed and corrected.

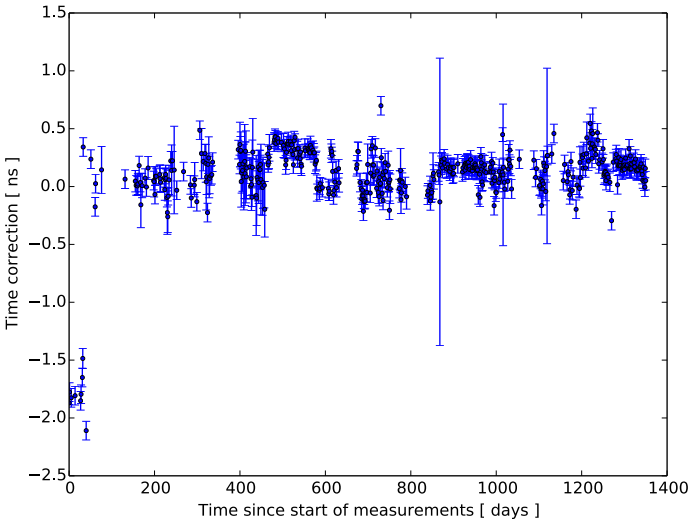


Figure 3.6: Time variation of the relative delay between two antennas within one LOFAR station over the course of our nearly four-year data collection. Residual delay values are binned per day, showing average and standard deviation within one bin.

3.4 Conclusion and outlook

We have developed a spectral cleaning method and a timing calibration method for interferometric radio antenna arrays. These were designed to operate on milliseconds-long time series datasets for individual receivers. The methods were used for our analysis of cosmic-ray datasets, to calibrate and clean voltage time series data. Using phases from an FFT for spectral cleaning was shown to be simpler to use than a straightforward threshold in an averaged power spectrum as no a priori knowledge of the antenna gain curve or noise spectrum is required. Moreover, when compared to this average spectrum threshold, the method has a slightly favorable detection power threshold that is at least 2.0 dB lower. In our application, the threshold of the method is at a power signal-to-noise ratio of −11 dB in a 25 kHz spectral window.

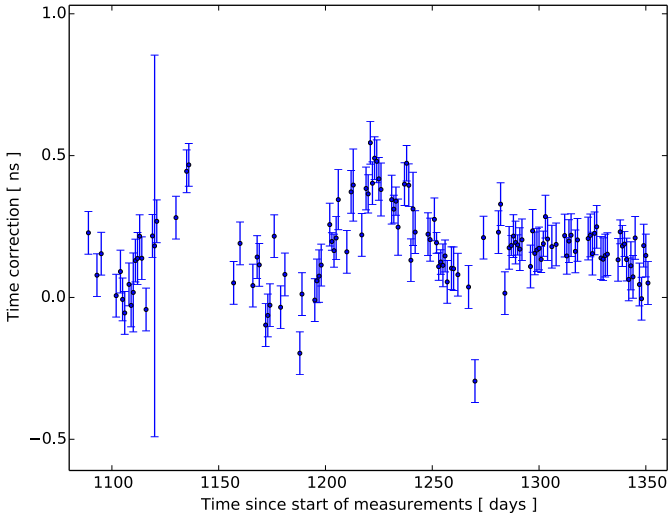


Figure 3.7: A close-up of the time variation of the relative delay for the same antenna pair, showing the precision of the delay monitoring as well as some clock drifting.

Timing calibration using the phases of public radio transmitter signals was performed to a precision of 0.4 ns for each antenna, at a sampling period of 5 ns (or 200 MHz sampling rate). Monitoring a given calibration over time has a precision of 0.08 ns for each antenna pair in a LOFAR station. Obtaining a timing calibration from a pulse transmitter aboard a drone flying over the array is possible to a similar precision of 0.3 ns, mainly limited by the accuracy of the position measurement of the transmitter.

As the methods described here only require datasets with lengths of 2 to 5 ms, they would be well suited for system monitoring and (pre-)calibration purposes of interferometric radio arrays in general. In addition to detecting interference and timing calibration, malfunctioning receiver data channels can also be identified. Examples include zero or unusual signal power, unstable timing calibrations, polarization errors, and outlying receiver gain curves. Detecting these issues at an early stage prevents the propagation of faulty signals into the correlation and imaging

process, where they are more difficult to remove.

It is expected that future low-frequency radio telescopes such as the SKA-Low (low-frequency part of the Square Kilometre Array) will also comprise many individual antenna elements laid out in a relatively dense pattern on the ground. In Dewdney (2015) it has been shown that the majority of antennas will be located at a distance of up to 10 km from a central core. These would be in the line of sight of a single transmitting beacon, either custom or RFI. Ideally a custom beacon would be used and turned on only a few parts per million of the time, for calibration.

Timing and phase calibration of all signal paths is a similar to the challenge in LOFAR, only on a much larger scale. Even with the use of one common clock signal, the entire signal path to the analog-digital conversion unit can exhibit nontrivial variations over time, e.g. along the analog signal transport to the central processing facility. This is already seen in Fig. 3.7, where the given antenna pair was located inside one LOFAR station, sharing the same clock signal. The techniques presented here, when merged with more elaborate existing methods, could prove useful for this.

Acknowledgements

The LOFAR Key Science Project Cosmic Rays gratefully acknowledges the scientific and technical support from ASTRON.

The authors thank K. Weidenhaupt, R. Krause, and M. Erdmann for providing and operating the octocopter drone. We also thank the anonymous referee for useful comments.

The project acknowledges funding from an Advanced Grant of the European Research Council (FP/2007-2013) / ERC Grant Agreement n. 227610. The project has also received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 640130). We furthermore acknowledge financial support from FOM, (FOM-project 12PR304) and NWO (VENI grant 639-041-130). AN is supported by the DFG (research fellowship NE 2031/1-1).

LOFAR, the Low Frequency Array designed and constructed by ASTRON, has facilities in several countries, that are owned by various parties (each with their own

funding sources), and that are collectively operated by the International LOFAR Telescope (ILT) foundation under a joint scientific policy.

3.A Appendix

Here we describe the details of the sensitivity analysis for the spectral cleaning method described in Sect. 3.2.1.

The problem of finding the threshold for detecting a transmitter can be reduced to a problem of determining whether a given random walk (or ensemble of random walks) is biased or not. The sum of a sequence of phase vectors $\mathrm{e}^{i\,\phi_j}$ forms a random walk in the complex plane, with unit step size. The random walk is biased if it has a preference towards a certain direction; on average this gives a longer distance for the random walk.

Assume a transmitter signal measured in one frequency channel of the FFT of a noisy time series, with amplitude a at each receiver. Let the mean noise power in this channel be σ^2, so the power signal-to-noise ratio is defined as $S^2 \equiv a^2/\sigma^2$. For this calculation, the receivers are assumed to have equal gain, which may not be the case in practice.

The noise in each frequency channel of an FFT is then Rayleigh-distributed in amplitude, with scale parameter $\sigma/\sqrt{2}$, and uniformly distributed in phase (Papoulis and Pillai, 2002). Therefore, denoting the random variable for the noise amplitude as b, the complex amplitude measured at two antennas can be written as

$$z_1 = a + b\,\mathrm{e}^{i\,\phi_1} \tag{3.8}$$

$$z_2 = a\,\mathrm{e}^{i\,\theta} + c\,\mathrm{e}^{i\,\phi_2}. \tag{3.9}$$

where $S^2 = a^2/\mathrm{E}(b^2)$. Here, $\mathrm{E}(\cdot)$ denotes expected value, and θ is the phase difference of the transmitter signal across the two antennas. As the noise phases are uniform-random and the following analysis is circular-symmetric, the transmitter signal phase difference θ can be omitted. For this analysis, the preferential direction of the random walk is then along the real axis.

Accumulating the phase variance $s_{1,2}$ for antenna indices 1 and 2 as in Eq. 3.3 corresponds to taking an average of the signal over all data blocks as follows:

$$s_{1,2} = 1 - \frac{1}{N_{\mathrm{blk}}} \left| \sum_{N_{\mathrm{blk}}} \frac{z_1\, z_2^*}{|z_1||z_2|} \right|. \tag{3.10}$$

As a first step, we calculate the expected value of the bias in the random walk. This follows from the expected value of the real part of the fraction in Eq. 3.10. As b and c are independent and identically Rayleigh-distributed, this expected value is given by

$$\frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} d\phi_1 d\phi_2 \int_0^{\infty} db\, dc \frac{2\,b}{\sigma^2} \mathrm{e}^{-b^2/\sigma^2}\, \frac{2\,c}{\sigma^2} \mathrm{e}^{-c^2/\sigma^2} \frac{\mathrm{Re}\left(a^2 + a\,c\,\mathrm{e}^{-i\,\phi_2} + a\,b\,\mathrm{e}^{i\phi_1} + b\,c\,\mathrm{e}^{i(\phi_1-\phi_2)}\right)}{\sqrt{a^2+b^2+2\,a\,b\cos(\phi_1)}\,\sqrt{a^2+c^2+2\,a\,c\cos(\phi_2)}}. \tag{3.11}$$

As we are dealing with low-amplitude thresholds well below the noise level (i.e. $S \ll 1$), an asymptotic lowest-order expansion in a/b is used in order to make the integral more tractable.

After collecting the lowest-order terms, the integral evaluates to

$$\mathrm{E}\left(\frac{\mathrm{Re}\,(z_1 z_2^*)}{|z_1||z_2|}\right) = \frac{\pi}{4}\, S^2 + \mathcal{O}(S^4). \tag{3.12}$$

The bias B in a random walk of N_{blk} steps is therefore expected to be

$$B = \frac{\pi}{4} S^2 N_{\mathrm{blk}}, \tag{3.13}$$

and the random walk effectively reduces, again to lowest order in S, to an unbiased random walk with respect to a point at distance B from the origin. Using the Rayleigh distribution for the unbiased random-walk distance to the origin, and displacing it by the bias, we obtain for the expected distance

$$\mathrm{E}(d) = \frac{1}{2\pi} \int_{-\pi}^{\pi} d\phi \int_0^{\infty} dR\, \frac{R}{\tau^2}\, \mathrm{e}^{-R^2/(2\tau^2)} \sqrt{R^2 + B^2 + 2\,R\,B\,\cos(\phi)}, \tag{3.14}$$

with scale parameter $\tau = \sqrt{N_{\mathrm{blk}}/2}$. To lowest order in B this yields

$$\mathrm{E}(d) \sim \mathrm{E}(d)_{\mathrm{unbiased}} + \frac{1}{4}\, B^2 \sqrt{\frac{\pi}{N_{\mathrm{blk}}}}. \tag{3.15}$$

The excess distance needs to be above a chosen factor k times the standard error of the unbiased random walk distance (see discussion of Eq. 3.4), using the ensemble having one random walk for each of the $N_{\rm ant}(N_{\rm ant}-1)/2$ antenna pairs. Hence we have a condition

$$\frac{1}{4}\,B^2\,\sqrt{\frac{\pi}{N_{\rm blk}}} > k\,\beta\,\sqrt{2}\,N_{\rm blk}^{1/2}\,N_{\rm ant}^{-1}, \tag{3.16}$$

where the right-hand side is k times the standard error for large $N_{\rm ant}$, approximating the number of antenna pairs by $N_{\rm ant}^2/2$.

Comparing these using Eq. 3.13 gives as a threshold for large $N_{\rm ant}$

$$S^2 > \frac{8}{\pi}\left(\frac{2}{\pi}-\frac{1}{2}\right)^{1/4}\,\sqrt{k}\,N_{\rm blk}^{-1/2}\,N_{\rm ant}^{-1/2}, \tag{3.17}$$

reducing to

$$S^2 > 3.8\,\sqrt{k/6}\,N_{\rm blk}^{-1/2}\,N_{\rm ant}^{-1/2}, \tag{3.18}$$

aimed at a six-sigma detection threshold ($k=6$).

Chapter 4

The effect of the atmospheric refractive index on the radio signal of extensive air showers

A. Corstanje et al. (LOFAR Key science project Cosmic Rays)

Astroparticle Physics 89, 23-29 (2017)

Abstract

For the interpretation of measurements of radio emission from extensive air showers, an important systematic uncertainty arises from natural variations of the atmospheric refractive index n. At a given altitude, the refractivity $N = 10^6\,(n-1)$ can have relative variations on the order of 10 % depending on temperature, humidity, and air pressure. Typical corrections to be applied to N are about 4 %. Using CoREAS simulations

of radio emission from air showers, we have evaluated the effect of varying N on measurements of the depth of shower maximum $X_{\max}$. For an observation band of 30 to 80 MHz, a difference of 4 % in refractivity gives rise to a systematic error in the inferred $X_{\max}$ between 3.5 and 11 g/cm^2, for proton showers with zenith angles ranging from 15 to 50 degrees. At higher frequencies, from 120 to 250 MHz, the offset ranges from 10 to 22 g/cm^2. These offsets were found to be proportional to the geometric distance to $X_{\max}$. We have compared the results to a simple model based on the Cherenkov angle. For the $120 - 250$ MHz band, the model is in qualitative agreement with the simulations. In typical circumstances, we find a slight decrease in $X_{\max}$ compared to the default refractivity treatment in CoREAS. While this is within commonly treated systematic uncertainties, accounting for it explicitly improves the accuracy of $X_{\max}$ measurements.

4.1 Introduction

In recent years, the techniques for measuring and modelling radio emission from air showers induced by high-energy cosmic rays have developed rapidly (Huege, 2016). The radio detection method has achieved high precision in estimating the air shower and primary particle properties (Buitink et al., 2014; Aab et al., 2016a; Bezyazeekov et al., 2016) which allows for very precise measurements of the primary particle type and the energy of the primary cosmic ray (Buitink et al., 2016). In particular, the LOFAR radio telescope (van Haarlem et al., 2013) has been used successfully for this, due to its densely instrumented core region located in the Netherlands. In an inner ring of 320 m diameter, we can use 288 low-band antennas, measuring in the $30 - 80$ MHz range, for cosmic-ray measurements. Additionally, there are also 288 high-band antennas measuring in the 110 to 190 MHz range. In an extended core region of about 6 km^2, nearly 1800 additional low-band antennas have been installed, grouped into stations of 96. Depending on strength and location of the air shower signal, up to four stations outside the inner ring can be used as well. The signals from air showers have been routinely measured with LOFAR since 2011 (Schellart et al., 2013).

In an air shower, secondary electrons and positrons are produced, which undergo charge separation as they travel through the Earth's magnetic field. This leads to transverse currents producing radio emission. This emission reaches the ground as a short pulse on the order of 10 to 100 ns long, with a specific lateral intensity distribution or 'footprint' (Nelles et al., 2015c) that depends on the depth of shower maximum X_{max}. The number of produced secondary particles peaks at X_{max}. This point, expressed as the column density of traversed matter ($\mathrm{g/cm^2}$), varies with primary particle type and is therefore an important quantity to measure in composition studies.

The measured lateral intensity distributions are compared to microscopic Monte Carlo simulations of air showers, to infer the properties of the primary cosmic ray. To simulate the radio signal at the antennas we use CoREAS (Huege et al., 2013), a simulation of the radio emission from the individual particles in the cascade simulated with CORSIKA (Heck et al., 1998). Fitting these simulated radio footprints to measured air showers allows us to infer X_{max} to a precision of $20\,\mathrm{g/cm^2}$. As this precision is comparable to that of fluorescence detection (Aab et al., 2014a; Kampert and Unger, 2012), it is well suited for composition studies. The radio detection method is therefore a complementary technique, as it is not limited to dark clear nights, and because its duty cycle is limited only by technical conditions and thunderstorms, it can reach in principle up to 100 %.

The detected radio signal depends on the difference in travel times of radio waves and particles. Therefore, it is important to apply an accurate value of the refractive index n. Variations in the refractive index lead to changes in the radio intensity footprint on the ground because the angle of peak emission depends on n. The refractive index, which at sea level is about $n \approx 1.0003$, exhibits natural variations due to weather conditions, at the level of the fifth decimal. It is therefore common to define the refractivity $N = 10^6\,(n-1)$, which emphasizes relative variations as these depend on $(n-1)$.

In this paper, we quantify the influence of the refractive index variations on the depth of shower maximum, to reduce the systematic uncertainty of the X_{max} measurements. In particular, CoREAS currently assumes a constant default value of the refractive index at each altitude in the atmosphere, and we explore how much the systematic error can be reduced by a more accurate treatment of the

local atmosphere. In the next section, we describe a toy model for the radiation from air showers, which qualitatively explains how unaccounted variations in the refractive index give rise to uncertainties in determining $X_{\max}$. In Sect. 4.3 we review the equations used to describe atmospheric parameters and their altitude profiles, and to calculate the refractive index. In Sect. 4.4, the method of fitting intensity distributions is described, and Sect. 4.5 gives the results for the systematic offsets of $X_{\max}$.

4.2 Toy model for the effect of the refractive index on radiation from air showers

The depth of shower maximum $X_{\max}$ can be inferred from the radio intensity footprint measured on the ground. In this section we show the radio footprint changes with the Cherenkov angle, which is a function of the refractive index. In particular, if the refractive index is higher than expected, the method based on the intensity footprint will underestimate $X_{\max}$.

In an extensive air shower, the magnetic field of the Earth induces an electric current, as the Lorentz force has opposite direction for electrons and positrons in the shower front. This current is transverse to the direction of the shower. The number of electrons and positrons depends strongly on the interaction depth of the shower and peaks at $X_{\max}$. The induced current is therefore strongly time dependent and emits electromagnetic radiation. The shower front has a thickness on the order of meters, and the radiation is coherent at wavelengths longer than this thickness, i.e. at radio wavelengths (Scholten et al., 2012). The transverse current resides in the shower front and thus moves towards the Earth surface with a velocity exceeding the speed of light in air (Werner et al., 2012). Therefore, radio waves are emitted because of coherent Cherenkov emission (de Vries et al., 2011). In addition to the emission from the transverse current there is also a smaller contribution from the net negative charge buildup in the shower front. From the polarization of the radio signal, this contribution was found to be 11 % on average at the LOFAR site (Schellart et al., 2014).

In the $30 - 80$ MHz band primarily used at LOFAR, the emission along the Cherenkov angle and the non-Cherenkov emission have roughly the same magnitude,

while at higher frequencies, the Cherenkov emission dominates (see Fig. 4 in de Vries et al., 2013). This has been confirmed by LOFAR observations above 110 MHz where a clear ring-like emission pattern is found (Nelles et al., 2015d). The radius of the ring was on the order of 100 m as is expected from the Cherenkov angle and the distance to $X_{\max}$. It has also been observed at GHz frequencies by the CROME (Smida et al., 2014) and ANITA (Hoover et al., 2010) experiments.

The angle α where the Cherenkov emission peaks, is given by

$$\cos\alpha = \frac{1}{\beta n}, \tag{4.1}$$

$$\alpha \approx \sqrt{2\,\beta(n-1)}, \tag{4.2}$$

where $\beta = v/c$ is the velocity of the shower front with $\beta = 1$ to a good approximation, the refractive index of air $n \approx 1.0003$ at sea level, and it varies with altitude. For convenience we also use the refractivity N throughout the text, as we will consider relative variations in N.

The depth of shower maximum $X_{\max}$ is reached at an altitude h_0, which is given by the altitude-dependent density $\rho(h)$ and the zenith angle θ. The relation is (Abreu et al., 2012)

$$X_{\max} \equiv X(h_0) = \frac{1}{\cos\theta}\int_{h_0}^{\infty} \rho(h)\,\mathrm{d}h, \tag{4.3}$$

where X is the column density expressed in g/cm^2. Therefore, X is also proportional to the pressure,

$$X(h_0) = \frac{10}{g}\frac{p(h_0)}{\cos\theta}, \tag{4.4}$$

with g the gravitational acceleration and p the pressure in Pa.

As a toy model for analyzing the effect of varying refractive index, we use the approximation that the size of the radio footprint is proportional to the base of a cone located in the shower plane, with a half-opening angle of α with respect to the the direction of the incoming primary particle. The shower plane is defined as the plane perpendicular to the incoming direction of the primary particle. Moreover, we assume that all radiation is produced near $X_{\max}$. As a consequence, the size of the radio footprint on the ground would be proportional to the geometric distance to $X_{\max}$ and to the Cherenkov angle at the altitude of $X_{\max}$. Variations in refractive index n at altitude h_0 would then translate to variations in radio footprint size via

the Cherenkov angle. The radio footprint with its non-circular symmetric structure (Huege et al., 2013; Nelles et al., 2015c) falls off smoothly with distance, hence it has no sharply defined “size". However, the important point for this model is that a given footprint would scale, both with distance to $X_{\max}$ and with the Cherenkov angle.

The model is expected to be more accurate at frequencies above 100 MHz where the Cherenkov mechanism dominates. At lower frequencies, the intensity pattern depends less strongly on the Cherenkov angle, and therefore on the refractive index.

As the refractive index is usually discussed in context of optical refraction, it should be noted that the additional effect of variations in n on bending of signal propagation paths (through Snell's law) is negligible for us. A signal path traveling from a medium with $n = 1$ to a medium with $n = 1.0003$ at a 60 degree incidence angle will be bent by 0.03 degrees. This is already below the resolution of about 0.1 degree attainable with LOFAR (Corstanje et al., 2015), and natural variations in n are still an order of magnitude smaller.

In composition studies, we fit radio intensity profiles simulated with CoREAS to the measured intensity profile on the ground (Buitink et al., 2014), to estimate $X_{\max}$. In Fig. 4.1, the effect of an increase in n is shown schematically, for a proton primary particle of 10^{17} eV from zenith. These have an average $X_{\max} \approx 670\,\mathrm{g/cm^2}$ (from CoREAS simulations), corresponding to an altitude of 3.51 km. If the refractive index is higher than expected, fitting the intensity profiles at ground will underestimate $X_{\max}$ (blue lines), as for the actual refractive index, the Cherenkov angle is larger (red lines). The lower altitude in the atmosphere corresponds to a higher $X_{\max}$; the modeled difference amounts to $17\,\mathrm{g/cm^2}$ for a 10 % increase in refractivity N. This is therefore a systematic uncertainty on $X_{\max}$.

The offset in $X_{\max}$ follows from the condition that the footprint size, which in this approximation is just the diameter of the intersection of the Cherenkov cone with the ground or shower plane, is kept constant. This gives

$$\sqrt{N_1/N_0}\, h_1\, \alpha(h_1) = h_0\, \alpha(h_0), \tag{4.5}$$

with h_0 and h_1 the altitude of the $X_{\max}$ point at fixed footprint size, for standard refractivity N_0 and $N_1 = 1.1\, N_0$ respectively. The square root arises from the small-angle approximation in Eq. 4.2; we have also taken $\sin\alpha \approx \alpha$. It holds both

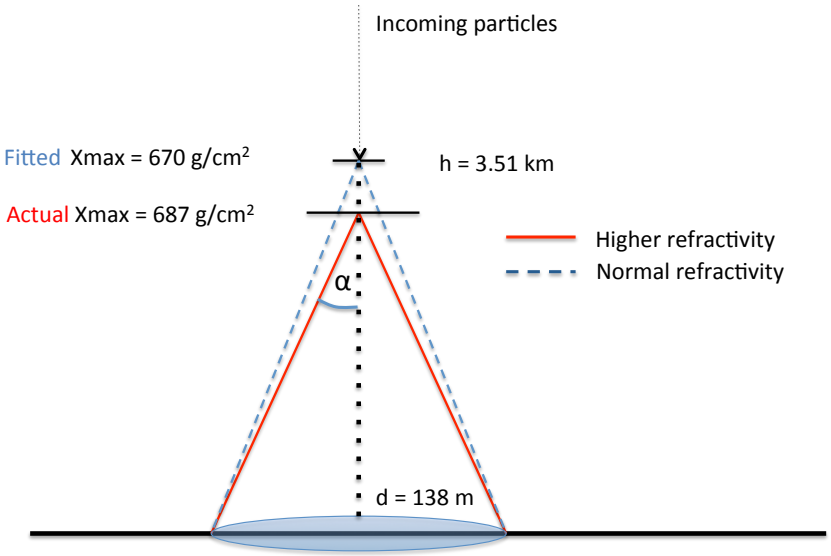


Figure 4.1: Schematic picture (stretched horizontally) of the effect of a higher refractivity on the estimate of $X_{\max}$. This model uses the Cherenkov angle α, and the fact that the radio emission is maximal around $X_{\max}$.

for vertical and for inclined showers and takes into account that N, and therefore α, varies with altitude as shown in the equations in the next section.

In summary, the steps for calculating the modelled shift in $X_{\max}$ given an original $X_{\max}$ at altitude h_0, are:

- establish N_0 at altitude h_0 using the equations in the next section (Eqs. 4.6 through 4.10).
- consider a different refractivity profile with altitude, e.g. $N_1(h) = 1.1\, N_0(h)$
- solve h_1 from Eq. 4.5
- obtain the column density (shifted $X_{\max}$) at this height, from Eq. 4.4 and Eq. 4.7 below

In the example of Fig. 4.1, we would have $h_0 = 3.51\,\mathrm{km}$, for which $N_0 = 192.3$. When considering a 10 % higher refractivity, i.e. $N_1(h) = 1.1\,N_0(h)$ at all altitudes, we obtain $h_1 = 3.31\,\mathrm{km}$ when keeping the footprint size constant. At this level, we would have a column density $X = 687\,\mathrm{g/cm^2}$, which is a systematic offset of $17\,\mathrm{g/cm^2}$.

4.3 The atmospheric refractive index

The refractive index n and refractivity N vary with temperature, pressure, and humidity in the atmosphere. As these depend on altitude, we use a parametrization for pressure and temperature corresponding to the US Standard atmosphere (U.S. Government Printing Office, 1976), where we have rewritten the equations to a slightly more compact form. We use only the bottom layer of this model, from 0 to $h = 11\,\mathrm{km}$, where $h = 0$ defines sea level. It is valid in the troposphere under 'average' circumstances,

$$T(h) = T_0 - Lh, \tag{4.6}$$

$$p(h) = p_0 \left(\frac{T(h)}{T_0} \right)^{\frac{gM}{LR}}. \tag{4.7}$$

Here, $p_0 = 1013.25\,\mathrm{hPa}$ and $T_0 = 288.15\,\mathrm{K}$ are standard sea level pressure and temperature. These can be varied according to local circumstances. The temperature lapse rate L is assumed constant at $L = 6.5\,\mathrm{K/km}$. The remaining constants are the gravitational acceleration $g = 9.80665\,\mathrm{m/s^2}$, the ideal gas constant $R = 8.31447\,\mathrm{J/(mol\,K)}$, and the molar mass of dry air $M = 0.0289644\,\mathrm{kg/mol}$. In the limit $L \to 0$, Eq. 4.7 reduces to the familiar exponential barometric formula.

Given the relative humidity H, the partial pressure of water vapor is calculated using the Magnus formula for the saturation pressure (Buck, 1981):

$$p_{\mathrm{sat}} = a \exp\left(\frac{b\,t}{t + c} \right) \tag{4.8}$$

$$p_w = H\,p_{\mathrm{sat}}, \tag{4.9}$$

where t is temperature in °C (in contrast to T in K), and p_{sat} is the saturation pressure for water vapor. For the empirically determined constants we take the values from Buck (1981), which are $a = 6.1121\,\mathrm{hPa}$, $b = 17.502$, $c = 240.97\,\mathrm{K}$ for

temperatures above 0 °C, and $a = 6.1115\,\mathrm{hPa}$, $b = 22.452$, $c = 272.55\,\mathrm{K}$ below 0 °C. The relative uncertainty in p_{sat} is then given as 0.1 % over the range of −50 to +40 °C .

The radio refractivity is parametrized according to Rüeger (2002) as

$$N = 77.6890\,\frac{p_d}{T} + 71.2952\,\frac{p_w}{T} + 375463\,\frac{p_w}{T^2}, \tag{4.10}$$

where T is temperature in K; p_d and p_w are the partial pressures (hPa) of dry air and of water vapor, respectively. The total air pressure is $p = p_w + p_d$. The small influence of carbon dioxide is included in the dry-air contribution. The accuracy of this equation is given as 0.02 % for the first term in Eq. 4.10, and 0.2 % for the second and third term combined. This evaluates to a relative uncertainty of less than 0.1 % in total, which is sufficient for our purposes.

It should be noted that at the radio frequencies of interest here, the refractivity values are different from those at infrared, visible, and UV wavelengths, such as considered in Abreu et al. (2012) for the fluorescence detection technique at Pierre Auger Observatory. In particular, the presence of water vapor significantly raises the radio refractivity, while it tends to lower the infrared refractivity slightly below that of dry air. The latter is depicted in Bernlöhr (2014) regarding optical Cherenkov telescopes. Therefore, the precision formulas of Edlén (updated in Birch and Downs, 1993) and Ciddor (Ciddor, 1996), defined for visible and near-infrared wavelengths are not applicable for us, and one can also not simply take the infinite-wavelength limit of those. For instance, for air at standard pressure, 20 °C, and 50 % relative humidity, the Ciddor equation gives $N = 268$, while Eq. 4.10 gives $N = 319$. It follows that for radio detection, accounting for humidity is more important than for fluorescence detection.

With the above formulas and definitions, we calculate the atmospheric profile of the refractivity versus altitude for different values of temperature, pressure, and humidity. In the US Standard atmosphere one uses the geopotential altitude, which takes into account the decrease in gravitational acceleration g with altitude. As the difference between geometric and geopotential altitudes at $h = 5\,\mathrm{km}$ is only 4 m, we do not correct for the difference.

An example plot of the altitude dependence of N is shown in Fig. 4.2 for two values of temperature and humidity. It is clear that humidity cannot be neglected,

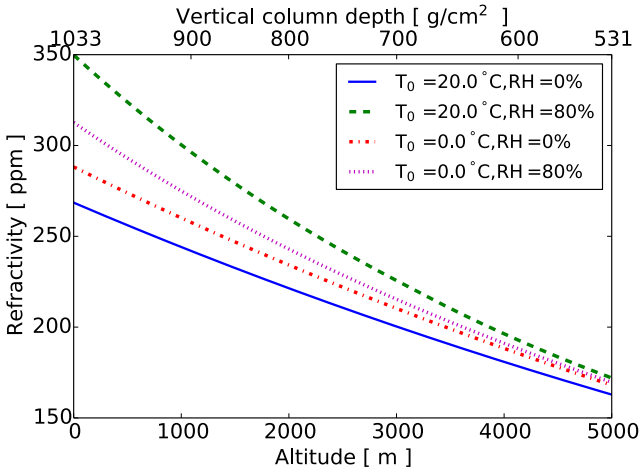


Figure 4.2: Four example altitude profiles of refractivity from Eqs. 4.10 and 4.6, all assuming standard pressure at sea level. For two values of sea-level temperature T_0, dry air is compared to a more realistic relative humidity (RH) in the Netherlands, around 80 %.

especially in the Netherlands where relative humidity near sea level is on average roughly 80 % (Royal Netherlands Meteorological Institute (KNMI), 2011). In CoREAS, the refractivity is set to a default value of $N = 292$ at sea level, and scaled with density according to the US Standard atmosphere, which considers dry air and $T = 15\,^{\circ}\mathrm{C}$ at sea level. In Fig. 4.3a, we show the relative correction factor to N with respect to its default value, at an altitude of 3.5 km. This corresponds to $X_{\max} = 670\,\mathrm{g/cm^2}$, which is the average value found for protons of 10^{17} eV energy. The correction is plotted against the ground temperature for several values of the relative humidity at the given altitude. At a sea-level temperature of 10 °C and 70 % relative humidity, the true refractivity is about 2 % lower than the default value. At higher sea-level temperatures, humidity plays a larger role, and the curves show a larger spread.

For air showers coming in at a 45 degree zenith angle, the altitude of $X_{\max}$ is about 6.1 km. At this altitude, the actual N is around 5 % lower than the default

value.

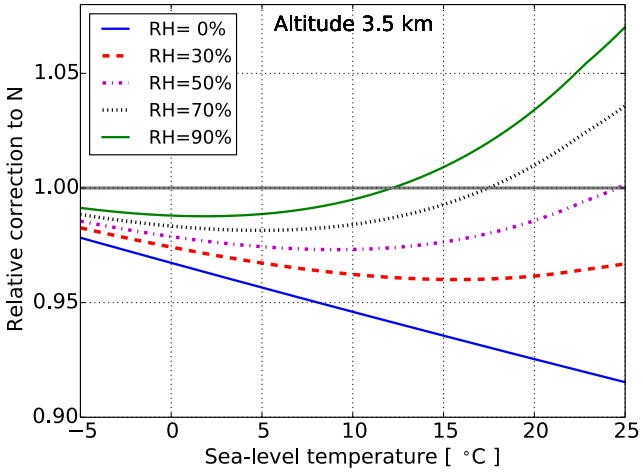
For the air pressure, we have taken the standard value of $p = 1013.25\,\mathrm{hPa}$. Natural variations, which are on the order of $\pm 2\,\%$, have the effect of lowering or raising the altitude of a given column density $X_{\max}$.

The relative humidity is expected to vary with altitude, dependent on conditions like cloud cover. The humidity in a region around the $X_{\max}$ altitude will be the most important. Therefore, to accurately estimate N, it is necessary to use atmospheric profile data such as available e.g. through the Global Data Assimilation System (GDAS) (National Oceanic and Atmospheric Administration, 2010). Uncertainties in these data translate into the uncertainty in N. Abreu et al. (2012) gives a comparison between GDAS data and weather balloon measurements in Argentina. Over the years 2009 and 2010, and in the altitude range of 3.5 to 6 km, this difference is $\pm 0.5\,^{\circ}\mathrm{C}$ for temperature, 0.5 hPa for pressure, and 0.05 hPa for water vapor pressure. The latter corresponds to about 2 % relative humidity at 3.5 km altitude, and about 7 % at 6 km.

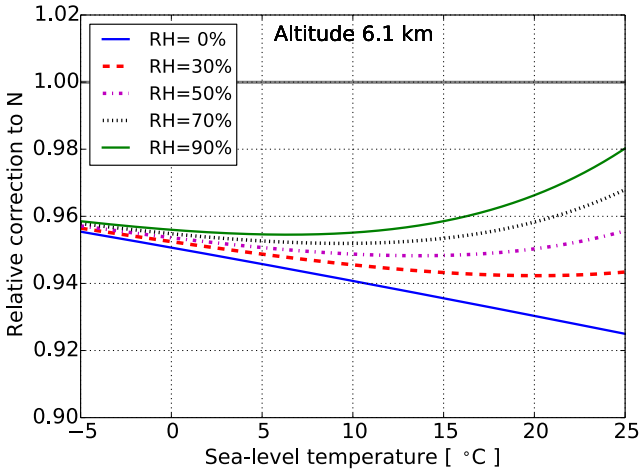
This shows that the GDAS data accurately represent the local circumstances. Although the uncertainty numbers may be different for other experiment sites, they are expected to be on the same order. The resulting relative uncertainty in N is around 0.5 % and is dominated by the humidity uncertainty. This is sufficient for our purposes.

4.4 Method

To investigate the effect of changing the refractive index on $X_{\max}$ measurements, we considered proton showers with a primary energy of $10^{17}\,\mathrm{eV}$ for five different zenith angles. For each case, we have generated an ensemble of 50 simulated showers with N at its default value of $N = 292$ at sea level. We have also generated another set of 50 showers where the refractivity N is higher by 10% at all altitudes. This is done using the same random number seeds, ensuring the evolution of the particles is exactly the same. Only the radio emission is recalculated, taking into account the higher refractivity. This method is similar to the one used in Buitink et al. (2014) that was used for the composition analysis at LOFAR (Buitink et al., 2016). The difference is that here we compare two simulated ensembles instead of comparing



(a)



(b)

Figure 4.3: Relative correction to the standard CoREAS refractivity, as a function of ground-level temperature and relative humidity at the given altitude. The altitudes corresponds the average $X_{\max}$ of a 10^{17} eV proton shower, (a) vertical, and (b) at a 45 degree zenith angle. In case (b), the correction profiles are always below unity in this temperature range.

simulations to measured data.

For all showers, we evaluate the lateral distribution of signal intensity in the shower plane $f(x, y)$, for a star-shaped pattern of antennas. In this pattern, 160 antennas are laid out along 4 lines, at a distance 25 m apart. The lines intersect at the origin and make angles of 45 degrees to each other, forming an octagonal star pattern. For the signal intensity we use a bandpass filter to limit the frequency range to $30 - 80$ MHz, relevant for the LOFAR cosmic-ray project. For comparison, we also consider the $120 - 250$ MHz band, where Cherenkov effects are more important.

We then take one of the showers with higher refractivity as a 'test shower', and fit the lateral distribution of each of the 49 other showers at the default refractivity to it. When fitting the shower with index k to the test shower, this yields a mean-square error as a fit quality measure:

$$\mathrm{MSE}(k) = \frac{1}{N_{\mathrm{ant}}} \sum_{\mathrm{antennas}} \left(A f_{\mathrm{test}}(x, y) - f_k(x, y) \right)^2 , \tag{4.11}$$

where $(x,\ y)$ are the antenna positions in the shower plane, and A is a scale factor that is taken as a free parameter. This is proportional to a reduced χ^2 for the case where all antennas have the same uncertainty on the intensity. Additionally, as we simulate antennas in a star-shaped pattern, we apply weight factors such that each antenna represents the same amount of area in the footprint. The optimal value of A that minimizes the MSE, is given by

$$A = \frac{\sum f_{\mathrm{test}}(x, y)\, f_k(x, y)}{\sum f_{\mathrm{test}}(x, y)^2}. \tag{4.12}$$

For every shower, we plot the fit quality (MSE) versus X_{max}. It is expected to have a minimum, and to lowest order, to have a quadratic dependence around the minimum. Therefore, we fit the points with a parabola. The location of the minimum of the parabola is used as an estimator for X_{max}. For one shower this is shown in Fig. 4.4, for a limited range around the minimum. The scatter around the fitted parabola arises from natural shower-to-shower fluctuations.

We found it useful to weight the fit-quality data points like shown in Fig. 4.4 by their inverse square in the parabolic fit, thus putting more emphasis on well-fitting profiles. This lowers the uncertainty in estimating the minimum X_{max} by up to 20 %, without introducing a bias. Proceeding this way, the inferred X_{max} of the

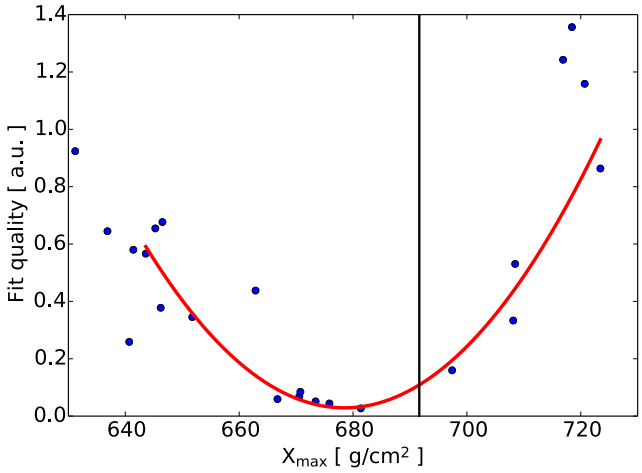


Figure 4.4: An example result of the fitting procedure of lateral distributions. This plot corresponds to one shower simulated at higher refractivity, fitted by an ensemble of showers at standard refractivity to evaluate the shift in $X_{\max}$. The simulated $X_{\max}$ is indicated by the black vertical line. Every dot corresponds to the fit to this shower of one shower from the ensemble simulated at normal refractivity. The fit qualities as defined in Eq. 4.11, when fitted by a parabola, show a minimum at $679\,\mathrm{g/cm^2}$, which can be compared to the originally simulated $X_{\max}$ value.

test shower with higher refractive index has a systematic offset with respect to its true $X_{\max}$ value. By taking in turn each of the 50 showers in the ensemble as test shower, this offset is evaluated along with its statistical uncertainty. This quantifies the effect of variations in the refractive index in e.g. a composition analysis where simulations are fitted to data.

We have limited the range of the fit to include those values of $X_{\max}$ within $\pm 40\,\mathrm{g/cm^2}$ of the expected value. As this range is determined by the true $X_{\max}$ and the offset, we shift the fit range in a second iteration. The offsets found in both iterations are consistent. We have discarded 5 showers at the low, and 5 at the high end of the true $X_{\max}$ range, as these give less accurate parabola fits due to lack of

data points at one side of the range.

The parabolic fit procedure has also been tested using only showers with the same refractive index. On average over 50 showers, the true $X_{\max}$ is reproduced within standard errors. For the ensemble of 50 minus 10 showers, the standard error ranges from 1 to 2 g/cm^2 for zenith angles from 15 to 50 degrees, respectively.

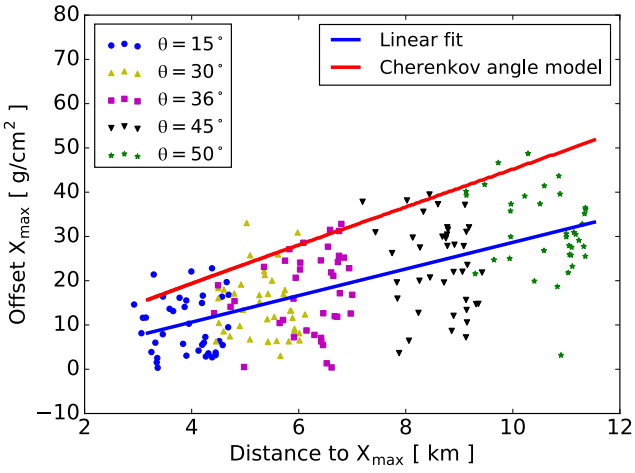
4.5 Results

We have evaluated the bias in $X_{\max}$ for a 10 % increase in refractivity, for five different zenith angles between 15 and 50 degrees. An increase in N, when not taken into account, leads to an overestimate in the geometric distance to $X_{\max}$, and hence to an underestimate in $X_{\max}$. Making the simplistic assumption that all radiation is produced at $X_{\max}$, and that the size of the radio footprint scales with the Cherenkov angle (Eq. 4.1), we apply the procedure given in Sect. 4.2 to find the shift in $X_{\max}$.

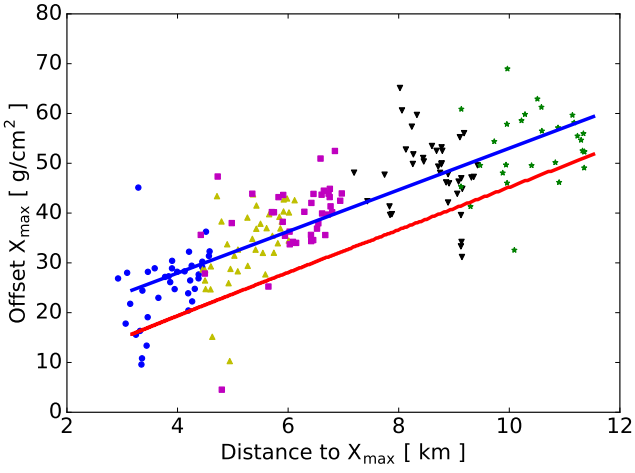
For the 30−80 MHz band, the results from the simulation fits are plotted together with the model predictions in Fig. 4.5a. A linear fit to the results for individual showers is also shown (blue line). The modeled offset in $X_{\max}$ is nearly linear in the distance to $X_{\max}$, and is a bit less than twice the offsets from simulations. The modeled offset reproduces the general dependence on geometric distance, but the simulated offsets are about a factor 2 smaller. As explained in Sect. 4.2, this is understood, as for low frequencies the non-Cherenkov emission is about equally important as the Cherenkov emission, leading to a weaker dependency on N. The offset ranges from about 9 g/cm^2 for a zenith angle of 15 degrees, to 28 g/cm^2 at $\theta = 50$ degrees. A linear fit suffices to describe the data; a higher-order polynomial curve does not significantly reduce the residuals. The standard deviation of the residuals is 8.9 g/cm^2, which is also the intrinsic uncertainty in this fit method, due to natural shower-to-shower fluctuations. The relation for the shift ΔX, defined as the underestimation of $X_{\max}$ per 10 % increase in refractivity, is

$$\Delta X_{10\,\%} = 3.00 \left(\frac{R}{1\,\mathrm{km}} \right) - 1.37\ \mathrm{g/cm^2}, \tag{4.13}$$

with R the geometric distance to $X_{\max}$. It can be scaled with the relative change in refractivity.



(a)



(b)

Figure 4.5: The $X_{\max}$ offset for zenith angles of 15, 30, 36, 45, and 50 degrees respectively, where the points represent individual simulated showers. The blue line is a linear fit to the data points. In (a), for the 30 to 80 MHz range, the simplistic model (red line) gives nearly twice the offsets found in simulations. For the 120 to 250 MHz range shown in (b), the simulated offsets are about 10 g/cm^2 below the simulations.

For the $120 - 250\,\mathrm{MHz}$ band, the $X_{\max}$ offset is shown in Fig. 4.5b. Here, the simulated offsets are well reproduced by the model, up to an almost constant additional shift. The fitted linear function for the offset is

$$\Delta X_{10\,\%} = 4.18 \left(\frac{R}{1\,\mathrm{km}} \right) + 11.24\ \mathrm{g/cm^2}. \tag{4.14}$$

The standard deviation of the fit residuals is $7.2\,\mathrm{g/cm^2}$ which is again not significantly reduced when using a quadratic function.

The relative correction to the refractivity from Fig. 4.3 is almost always below unity for vertical showers, and always below unity for inclined showers in the temperature range relevant at LOFAR. Therefore, this gives a systematic offset, not only in individual $X_{\max}$ estimates, but also in the average $X_{\max}$ inferred from many air showers.

Over the range -5 to $20\,^{\circ}\mathrm{C}$, the actual N is about $1.5\,\%$ below the default simulated value for vertical showers. For inclined showers, the actual N is about $4.5\,\%$ below the default. Hence, the average $X_{\max}$ is overestimated by about $1.5\,\mathrm{g/cm^2}$ for near-vertical showers, and by about $11\,\mathrm{g/cm^2}$ at 45 degrees inclination.

The bias is larger for more inclined air showers; it has been noted in the LOFAR composition study (Buitink et al., 2016) that more inclined air showers (above 32 degrees zenith angle) had on average a higher inferred $X_{\max}$ than the more vertical ones. The difference in average $X_{\max}$ amounts to $16\,\mathrm{g/cm^2}$ between the two cases. From the offsets in Eq. 4.13, it follows that at most $9\,\mathrm{g/cm^2}$ of this can be explained by a bias caused by using an incorrect value of the refractive index.

The fact that N is slightly overestimated in simulations is also relevant for other radio detection experiments such as AERA (Aab et al., 2016a) and Tunka-Rex (Bezyazeekov et al., 2016). These experiments, located in Argentina and Siberia respectively, have a temperature range different from the LOFAR site. Nevertheless, for the case of inclined showers, the relative correction as from Fig. 4.3b is below unity down to well below $-40\,^{\circ}\mathrm{C}$.

The numerical constants in Eqs. 4.13 and 4.14 follow from the simulations, which depend on location parameters such as height above sea level and the geomagnetic field vector. Therefore, for other experiments the bias on the average $X_{\max}$ is expected to be on the same order, but with slight differences due to variations in location-specific parameters, leading to different constants in Eqs. 4.13 and 4.14.

4.6 Summary

The technique of measuring radio signals from air showers to infer the mass composition of cosmic rays, relies on accurate measurements of the depth of shower maximum $X_{\max}$. One of the systematic uncertainties on $X_{\max}$ is given by the refractive index of air in the atmosphere, which exhibits natural variations.

We have evaluated the effect of variations in the refractive index of air on determining $X_{\max}$. Using a procedure similar to that used in the composition study at LOFAR (Buitink et al., 2016), we have taken simulated proton showers at a 10 % higher refractivity $N = 10^6\,(n-1)$. We have fitted them with a Monte Carlo ensemble of 49 showers simulated at a default value of N, excluding the one corresponding to the higher-N shower being fitted. The minimum in the least-squares fit quality yields, on average over many showers, the systematic offset in $X_{\max}$.

These offsets were found to be proportional to the geometric distance to $X_{\max}$, for zenith angles ranging from 15 to 50 degrees. The effect is roughly twice as strong for the $120 - 250\,\mathrm{MHz}$ band as for the $30 - 80\,\mathrm{MHz}$ band. Given variations in N on the order of 4 %, from the atmospheric effects described in Sect. 4.3, the offsets would range from about 3.5 to $11\,\mathrm{g/cm^2}$ for $30 - 80\,\mathrm{MHz}$, and from 10 to $22\,\mathrm{g/cm^2}$ for $120 - 250\,\mathrm{MHz}$.

A simplistic model in which the radio intensity pattern is assumed to be proportional in size to the Cherenkov cone starting from the $X_{\max}$ point, qualitatively describes the effect; the fitted offsets were found to be just above half the modeled offsets in the $30 - 80\,\mathrm{MHz}$ band, and about 20 % above the modeled offsets in the $120 - 250\,\mathrm{MHz}$ band.

Calculated profiles of refractivity versus altitude show that one cannot use a single default value of N in simulations. This leads to a bias in the average $X_{\max}$, depending on the choice of a constant N either for near-vertical or for inclined showers. The accuracy can be improved for individual showers by using Eqs. 4.13 and 4.14.

As a further improvement, it would be required to include detailed atmospheric data, including the particular refractivity profile at the time of the air shower, into a next version of Corsika / CoREAS. The GDAS database (National Oceanic and Atmospheric Administration, 2010) is useful for this. This allows to fully account

for the refractive index variations, and to re-evaluate the LOFAR measurements at the best level of detail. This is the subject of a future publication.

Acknowledgements

We acknowledge financial support from the Netherlands Organization for Scientific Research (NWO), VENI grant 639-041-130, the Netherlands Research School for Astronomy (NOVA), the Samenwerkingsverband Noord-Nederland (SNN) and the Foundation for Fundamental Research on Matter (FOM). We acknowledge funding from the European Research Council under the European Union's Seventh Framework Program (FP/2007-2013) / ERC (grant agreement n. 227610) and under the European Union's Horizon 2020 research and innovation programme (grant agreement n. 640130).

Chapter 5

Measuring the composition of cosmic rays from 10^{17} to 10^{18} eV with the LOFAR radio telescope

A. Corstanje et al. (LOFAR Key science project Cosmic Rays)

in preparation

5.1 Introduction

Cosmic rays arrive at the Earth's atmosphere in an energy range from 10^9 to over 10^{20} eV. Upon interacting in the atmosphere, they produce a cascade of secondary particles called *extensive air shower*, which is measurable in ground-based detector arrays for energies above about 10^{14} eV. At the high end of the energy spectrum, these particles have the highest energy of the known particles in the universe, and

their interactions in the atmosphere occur at considerably higher center-of-mass energy than those in man-made accelerators on Earth. Therefore, the questions about their origin and their mass composition have raised considerable interest, and cosmic-ray air showers are measured in observatories around the world. The largest is the Pierre Auger Observatory in Argentina (Abraham et al., 2010, 2004), spanning an area of $3000\,\mathrm{km}^2$.

In this analysis, we study cosmic rays with a primary energy between 10^{17} and $10^{18}\,\mathrm{eV}$, the energy range where a transition takes place from particles originating from within the Galaxy, to an extragalactic origin. From the Hillas criterion (Hillas, 1984), there is a maximum energy for particles to be magnetically contained in a region of a given size, such as the Galaxy. This maximum energy is proportional to the charge of the particles. Hence at some point in the given energy regime, it is expected that heavy nuclei, i.e. carbon and heavier, are still mainly from Galactic origin, while light nuclei, i.e. protons and helium, are already mainly extragalactic (see e.g. Blümer et al., 2009). Therefore, composition measurements in this energy region are interesting for comparison with models of cosmic-ray sources and propagation.

Along the track of an air shower, the number of secondary particles rises, reaches a maximum, and then falls again. The position of the maximum number of charged particles, expressed in $\mathrm{g/cm}^2$ of traversed matter, is referred to as the depth of shower maximum, X_{max}. The atmosphere is thick enough for this maximum to be reached for almost all showers, typically at altitudes of 2 to 7 km depending on zenith angle, energy, and mass of the primary cosmic ray.

At a given primary energy, X_{max} depends on the mass of the primary particle. On average, it is different for protons compared to heavy nuclei, and also the shape of the X_{max}-distribution differs. Protons penetrate deeper on average (higher X_{max}), with a wide distribution around the average. Showers from iron nuclei have a narrower distribution around a lower average X_{max}. The shift in average X_{max} with respect to protons is approximately proportional to $\ln A$, for particles with atomic mass number A. Thus, measuring X_{max} for a collection of air showers gives information about their composition.

There are four main techniques for measuring X_{max}: (i) measuring the secondary particles on the ground, especially the electron/muon ratio, (ii) measuring fluores-

cence light along the trail of the air shower, (iii) measuring Cherenkov light, and (iv) measuring the radio signal using antennas on the ground (Kampert and Unger, 2012; Huege, 2016). The radio detection technique has shown substantial development in recent years, leading to a method to determine $X_{\max}$ with a resolution about $20\,\mathrm{g/cm^2}$ (Buitink et al., 2014). The method has been demonstrated using the LOFAR radio telescope, showing that the cosmic rays around $10^{17}\,\mathrm{eV}$ have a considerable light-mass component (Buitink et al., 2016). Here, we present a method which has been improved on several points, thus lowering the systematic uncertainties.

The method relies on air shower simulations tracking individual particles, and summing up their contributions to the radio signal measured on the ground. For this, the CORSIKA (Heck et al., 1998) simulation program has been used, with its plugin CoREAS (Huege et al., 2013) for computing the radio signal. For an ensemble of simulated air showers, their lateral intensity distribution or 'radio footprint' is fitted to the measurements, from which $X_{\max}$ and the energy of the measured shower are reconstructed.

One of the main systematic uncertainties in the $X_{\max}$ measurements is given by the local atmospheric conditions at the time of the air shower. The atmospheric density profile has a geometric effect, through the conversion from $X_{\max}$ in $\mathrm{g/cm^2}$ to a distance from the detector. A first-order correction for this effect was already applied in Buitink et al. (2014).

In Corstanje et al. (2017) it was shown that accurately representing the refractive index in the atmosphere, i.e. its altitude profile at the time of the air shower, is important for precision measurements; not accounting for its natural variations leads to a systematic error of 4 to $11\,\mathrm{g/cm^2}$ given typical weather variations at the LOFAR site. In the present analysis, we have included the local atmospheric variations into our air shower simulations, by representing the altitude profiles of density and refractive index. As input, we use the atmospheric data from the Global Data Assimilation System (GDAS) (National Oceanic and Atmospheric Administration, 2010).

Other improvements to the analysis include a radio-only reconstruction of both $X_{\max}$ and energy, the latter using a new calibration based on Galactic emission (Mulrey et al., 2019). Using a fast pre-computation of shower simulations with CONEX (Bergmann et al., 2007) streamlines the reconstruction, as showers can be

pre-selected for their $X_{\max}$. The selection criteria to obtain a bias-free $X_{\max}$ sample have been improved, and a refined statistical analysis has been done. All these increase the accuracy of the composition analysis, by lowering systematic and/or statistical uncertainties.

The Low Frequency Array (LOFAR) (van Haarlem et al., 2013) is a radio telescope consisting of many separate antennas. The core region in the north of the Netherlands has a high density of antennas. The antennas are grouped in stations, each of which in the Netherlands contains 96 low-band antennas (LBA), working in the 10 to 90 MHz range, and 48 high-band antennas (HBA) operating at $110 - 240$ MHz. The center of LOFAR is a circular area of 320 m diameter, with six of those stations. In a core region of about 6 km^2, there are 18 more stations. LOFAR uses ring buffers to store up to 5 seconds of the raw measured signals at each antenna, which are used to measure the radio signals of air showers.

To trigger a buffer readout when an air shower arrives, a particle detector array called LORA (LOFAR Radboud Air shower Array) (Thoudam et al., 2014) is located inside the innermost ring of LOFAR. With 20 scintillator detectors monitored in real time, a trigger is sent to LOFAR when a threshold of 13 coincident detections is reached, a level which is optimal for our purposes.

The paper is organized as follows: in Sect. 5.2, we present the method of fitting air shower simulations to measured data to infer $X_{\max}$. Furthermore, in Sect. 5.2.4 we discuss the selection criteria used to obtain a bias-free sample of showers. In Sect. 5.3, the statistical analysis to infer particle composition from the $X_{\max}$ values is explained. The results are split into two sections, Sect. 5.4 for the $X_{\max}$ distribution from our dataset, and Sect. 5.5 for the composition results. A summary is given in Sect. 5.6.

5.2 Method

The discussion of the methods is split into three sections: first, we give a brief review of the procedure to infer $X_{\max}$ for individual measured air showers. A more detailed explanation is found in Buitink et al. (2014); the details that have changed in this version are given below. Second, we show how including the local atmospheric conditions into the simulations leads to improved accuracy. Finally, we explain our

method to select showers in order to create an unbiased sample.

5.2.1 Using CoREAS simulations to estimate $X_{\max}$ of measured air showers

For each shower measured with LOFAR, we produce an ensemble of CoREAS showers, spanning the natural range of $X_{\max}$ for protons and for iron nuclei. We use the QGSJetII-04 hadronic interaction model (Ostapchenko, 2013) to produce the particle showers with CORSIKA. As simulation energy we use a pre-estimate from fitting an analytic description of the radio footprint (Nelles et al., 2015c), or an estimate from the particle detectors when the fit failed to converge.

As a pre-computation stage we produce 600 showers with the fast, approximate simulation method CONEX (Bergmann et al., 2007). This ensemble is used to select those random number seeds to span the natural range of $X_{\max}$ roughly uniformly with about 15 showers. The same random number seeds are used in the full CORSIKA simulations. The aim is twofold, to have simulated showers covering the entire range of $X_{\max}$, which is important for the selection criteria for a bias-free sample (see Sect. 5.2.4), and to have a region around the best-fitting $X_{\max}$ with extra dense coverage, to improve precision.

Therefore, ten additional showers are simulated in a region of $\pm 20\,\mathrm{g/cm^2}$ around the first $X_{\max}$ estimate, aiming to have a high density of simulations close to the reconstructed $X_{\max}$. If the reconstructed $X_{\max}$ deviates from the initial fit, extra showers are simulated to match the dense region with the reconstructed $X_{\max}$. An example is shown in the middle panel of Fig. 5.1; existing showers with $X_{\max} > 700$ fall outside the plotted vertical range.

The radio signal of each simulated shower is passed through our antenna model for the LOFAR LBA antennas (Schellart et al., 2013), and through the bandpass filter used in the data analysis, to be able to compare with LOFAR data. The signal intensity for each simulated shower is then matched per antenna to the LOFAR measurements. In this fit, the core position and an overall scaling factor are free parameters. This gives a chi-squared value for each shower:

$$\chi^2_{\mathrm{radio}} = \sum_{\mathrm{antennas}} \left(\frac{P_{\mathrm{ant}} - f_r^2\, P_{\mathrm{sim}}\left(x_{\mathrm{ant}} - x_0,\, y_{\mathrm{ant}} - y_0\right)}{\sigma_{\mathrm{ant}}} \right)^2, \tag{5.1}$$

where P_{ant} and σ_{ant} denote the measured signal intensity and its uncertainty, and P_{sim} is the simulated intensity. The overall scaling factor is f_r^2, and (x_0, y_0) is the fitted shower core position. In contrast to the method in Buitink et al. (2014), we perform the fit based on the radio signals only, making the radio reconstruction self-sufficient. Before, the fit included both radio and particle detector data. As a consequence, showers for which the reconstruction cannot be done accurately without the particle detector signals are now (automatically) discarded.

The result of the fitting procedure for one of our measured showers is shown in Fig. 5.1. In the left panel, the best-fitting simulated shower is shown (background color) together with the measurements (colored circles). The colored circles blend in well with the background color, indicating a good fit. This is confirmed by the middle plot, showing a reduced χ^2 of 1.3 for the best fit, and a clear minimum as a function of X_{max}. The right panel shows a one-dimensional representation of the simulated and measured intensities per antenna.

We employ a Monte Carlo procedure to infer the uncertainties on X_{max}, the energy and the shower core position. For each simulated shower in our ensemble we create 3 mock datasets as they would have been measured, i.e. adding the noise level found in each LOFAR antenna. They represent 3 different realisations of the random noise, at a fixed shower core position. This simulated shower is then reconstructed through the above procedure, using the ensemble of all other simulations. Statistically comparing the reconstructions with the real X_{max}, core position, and radio scale factor, which are known in simulations, yields their uncertainties. The uncertainties thus found are calculated from the entire simulated ensemble, and are applicable to the measured shower as well as to each simulated shower; this is important in the bias-free sample selection procedure explained in Sect. 5.2.4.

This procedure, relying on the reconstruction method described above, now also uses only the radio signals.

5.2.2 Measurement of the primary particle energy

We measure the energy of the primary particle by comparing the intensity of the measured radio signal with the predicted radio signal from CoREAS, which was produced at a given simulation energy obtained from the initial fit. The intensity of the radio signal scales quadratically with the primary energy (Zimmermann et al.,

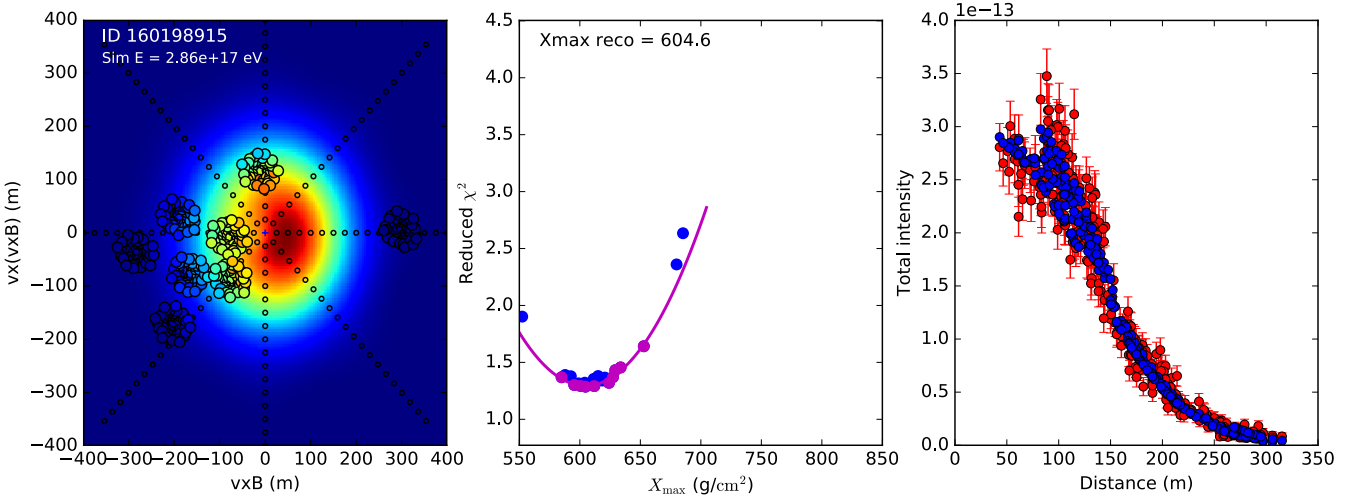


FIGURE 5.1: An example of fitted CoREAS showers to a LOFAR-measured shower. The left panel shows simulated signal intensity along with the measurements, in the shower plane, for the best-fitting shower. The lateral intensity distribution is not rotationally symmetric. The middle panel shows the reduced χ^2 as a function of $X_{\max}$, with a parabolic fit through the lower envelope denoted by the magenta line. The right panel shows a 1-D lateral distribution function, where red points with margins denote the measurements, and blue points denote the simulated intensities.

2017; Nelles et al., 2015b). Fitting CoREAS radio intensity to LOFAR data produces an overall scale factor. The square root of this is taken as a correction to the simulation energy, giving an estimate of the primary particle energy.

This procedure of matching simulated and measured intensity relies on an accurate absolute calibration of the radio antennas at LOFAR. The calibration has been improved recently, using the emission from the Galaxy and taking into account various noise contributions due to the electronics (Mulrey et al., 2019). The calibration has a systematic uncertainty of 13 % in amplitude, which is therefore the contribution of the calibration to the systematic uncertainty in energy.

As a cross-check, we have also estimated the energy based on the signals from the particle detectors. To this end, one compares the particle footprint from the best-fitting Corsika shower to the signals at the LORA detectors. The conversion of particles reaching the ground to LORA signals is simulated using the GEANT4 simulation package (Agostinelli et al., 2003), a procedure described further in Sect. 5.2.4.1. We find an agreement within 10 % between the resulting energy scales from radio and from particles.

The statistical uncertainty on the energy estimate follows from our Monte Carlo uncertainty analysis per shower (see Sect. 5.2.1). The average values for the uncertainties on energy and log-energy, i.e. σ_E and $\sigma_{\log E}$ correspond to 9 and 10 %, respectively. This is a considerable improvement over the 32 % uncertainty in Buitink et al. (2016), and arises from the large number of radio antennas compared to the 20 particle detectors used earlier.

The systematic uncertainty on the energy is taken as 27 %, as from the previous analysis based on the particle detectors (Buitink et al., 2016).

5.2.3 Improved accuracy by including local atmospheric parameters

To improve the accuracy of the simulations and the $X_{\max}$ reconstructions, we have now used an updated version of CORSIKA and CoREAS (v7.6300), which allows to include the local atmospheric conditions into the simulation runs.

The atmospheric parameters at the time of each air shower are taken from the Global Data Assimilation System (GDAS) (National Oceanic and Atmospheric Administration, 2010), which gives pressure, temperature, and humidity in 24 layers

in the atmosphere. These are data used e.g. in weather models.

The important quantities for us are the altitude profiles of density and refractive index. The density profile determines the amount of matter traversed by the particles along the shower evolution, and therefore the shower geometry depends on this. Due to natural variations in air pressure and temperature, the geometric distance to $X_{\max}$ may be under- or overestimated, leading to a systematic error per shower on the order of 15 to $20\,\mathrm{g/cm^2}$. In the earlier analysis of Buitink et al. (2014), the GDAS density profile was used to correct to first order the $X_{\max}$ estimate from simulations using the US Standard Atmosphere.

The refractive index n and its variations are important for the radio emission processes, even though at sea level $n \approx 1.0003$ is only slightly above unity. The refractive index is a function of both the density and the humidity. Natural variations in n make the Cherenkov angle wider or narrower, thus affecting the intensity footprint on the ground (Corstanje et al., 2017). Typical variations of $(n-1)$ are on the order of 4 %, and introduce a systematic error on the inferred $X_{\max}$. Using CoREAS simulations, this error was found to be about 4 to $11\,\mathrm{g/cm^2}$, depending on the zenith angle, with the higher error at a zenith angle of 50 degrees.

Residual errors in $(n-1)$ as taken from GDAS temperature, pressure, and humidity are about 0.5 %. Hence, errors on $X_{\max}$ will be on the order of 1 to $2\,\mathrm{g/cm^2}$, and will vary between positive and negative from one shower to another, adding to the statistical uncertainty per shower. We have thus removed a systematic uncertainty that is important for precision measurements.

In CORSIKA, five layers are used to parametrize the atmospheric density profile as a function of altitude. In each layer, the density is set to fall off exponentially with altitude, with a scale height as a free parameter. We have used least-squares curve fitting to determine the optimal parameters to match the 5-layer model atmosphere to the GDAS representation (Mitra et al., 2019). Boundary conditions for the fit are given by the requirement of continuity of the density at the layer boundaries. The error on $X_{\max}$ induced by the 5-layer approximation was found to be about $4\,\mathrm{g/cm^2}$, and adds to the statistical uncertainty per shower. It introduces a systematic uncertainty of 1 to $2\,\mathrm{g/cm^2}$, depending on altitude, hence taken as $2\,\mathrm{g/cm^2}$.

5.2.4 Bias-free sample selection

In this section we show how to apply fiducial cuts, i.e. to reject showers that would introduce a composition bias to the sample. For the composition measurement, we aim to obtain a sample which is unbiased in $X_{\max}$. We do not expect, however, to obtain a sample reflecting the natural cosmic-ray energy spectrum, as the effective exposure area, both on the ground and on the sky, depends strongly on energy.

We analyze each measured shower given its energy, reconstructed shower core position, and incoming direction. The central requirement is that this shower would have produced a trigger in both the particle detectors and in the radio data, would it have had any other value of $X_{\max}$ in the natural range. Moreover, it must meet the core reconstruction quality criterion explained below.

A dataset comprising all measured showers that meet this requirement is then unbiased in $X_{\max}$, i.e. it is a sufficient condition. Due to the irregular array layout and moderate event count, a per-shower inclusion criterion is more efficient than attempting to construct a fiducial volume in parameter space (which would also be rather irregular).

A bias may arise from the particle detector trigger, which is reached more easily for showers penetrating deeper into the atmosphere (high $X_{\max}$). Another, opposite source of bias arises from the radio detection threshold. We require at least 3 LOFAR stations to detect significant pulses for a given shower. Showers with low $X_{\max}$ have a larger radio footprint, and hence are more likely to trigger three LOFAR stations.

5.2.4.1 Removing bias from the particle trigger

For each measured shower, we use the set of all simulated showers, including their particle content, to see if each simulated shower would trigger LORA. For this, we use the GEANT4 simulation tool (Agostinelli et al., 2003), which simulates the particles traversing the detectors and their deposited energy. The simulation of the LORA detectors was also used in the measurement of the cosmic-ray energy spectrum presented in Thoudam et al. (2016). Only if all showers in the ensemble would trigger, the measured shower is included in the sample.

From CORSIKA we obtain a list of particles reaching the ground, with their respective positions and momenta. In the GEANT4 simulation, this is converted to an energy deposit at the detector locations. We divide the energy deposit by an

average value of 6.7 MeV per particle. The value of 6.7 MeV arises from the most probable energy deposit of single, high-energy muons from an all-sky distribution (Thoudam et al., 2016). Although the muons vary in energy, their deposit is nearly constant with energy, and a detector is triggered by a single muon.

When one or more particles hit a detector, it will trigger; as this is subject to Poisson statistics, we evaluate the probability of zero particles at the given energy density. Out of 20 LORA detectors, 13 must trigger in coincidence for the radio data of the air shower to be recorded. Hence, also from all simulated showers we require that with a probability of 99 %, at least 13 detectors would trigger (due to statistical fluctuations, the probability cannot reach exactly 100 %).

This test has a tendency to remove showers from the sample which have large reconstructed $X_{\max}$ values, i.e. relatively low in the atmosphere, and/or high inclination. In this case, the given measured shower has produced a trigger, but had its $X_{\max}$ been lower, the number of particles would have been too small. Similarly, showers with low energy and/or a core position far from the LORA detectors are more likely to be rejected.

5.2.4.2 Removing bias from the radio detection threshold

We perform a test against bias from the finite radio detection threshold. The criterion is, similar to the particle detection bias test, that the radio signal for each simulated shower in the ensemble would have been detected above the noise in at least three LOFAR stations.

To this end, we take the core position of the shower that fits best to the LOFAR-measured shower, and position also all other simulated showers here with respect to LOFAR. From the best-fitting shower, we have a fitted scale factor relating simulated intensity to measured intensity. Using this scale factor, we obtain the pulse intensities for each simulated shower and for each antenna. The noise intensities from the LOFAR-measured showers are taken as reference, and a threshold criterion is set as an intensity signal-to-noise ratio of 6 in each antenna. In the data processing pipeline, half of the antennas per station are required to trigger to have a 'good' detection. Although that detection is amplitude-based, an intensity signal-to-noise ratio of 6 was found to be slightly conservative, and otherwise in good agreement with the amplitude threshold detection.

This test typically rejects showers with a small reconstructed $X_{\max}$ value, i.e. relatively high in the atmosphere, and/or zenith angle; a shower with the same parameters would then have a much smaller radio footprint at high $X_{\max}$, which may not be able to trigger three LOFAR stations.

5.2.5 Reconstruction quality cuts

The procedure described in Sect. 5.2.1 to infer the uncertainty on $X_{\max}$ is also useful as a test of the reconstruction quality of the radio signal. Apart from the $X_{\max}$ uncertainty, it also gives an uncertainty on the fitted shower core position and on the energy. These uncertainties are calculated from the entire simulated ensemble, and hence they are the same for each simulated shower being tested by the two above procedures.

From the three uncertainties, the precision of the core position reconstruction is arguably the most relevant indicator of overall shower reconstruction quality. When this precision is low, one cannot expect either $X_{\max}$ or energy to be reconstructed accurately. Shown in Fig. 5.2 is the uncertainty on $X_{\max}$ versus the core position uncertainty. They are clearly correlated, and a cut on the reconstruction uncertainty at 7.5 meters was found to be sufficient to reject the majority of poorly reconstructed showers, while retaining showers with low $X_{\max}$ uncertainty.

The appearance of poorly reconstructed showers, despite meeting the other criteria, comes mainly from the position of some showers with respect to the LOFAR array geometry. Most notably, when the core position is outside the array and/or only three stations have been triggered at low signal-to-noise ratio, the reconstruction precision becomes well below average. This criterion catches these cases automatically.

5.2.6 Systematic uncertainties

Our method to determine $X_{\max}$ is affected by the following systematic uncertainties, which are summarized in Table 5.1. The choice of the hadronic interaction model used in CORSIKA, in this case QGSJetII-04, introduces a systematic uncertainty of $5\,\mathrm{g/cm^2}$ (Buitink et al., 2016) in the $X_{\max}$ measurements, due to minor differences in radio footprints when changing the model e.g. to EPOS-LHC (Pierog et al., 2015).

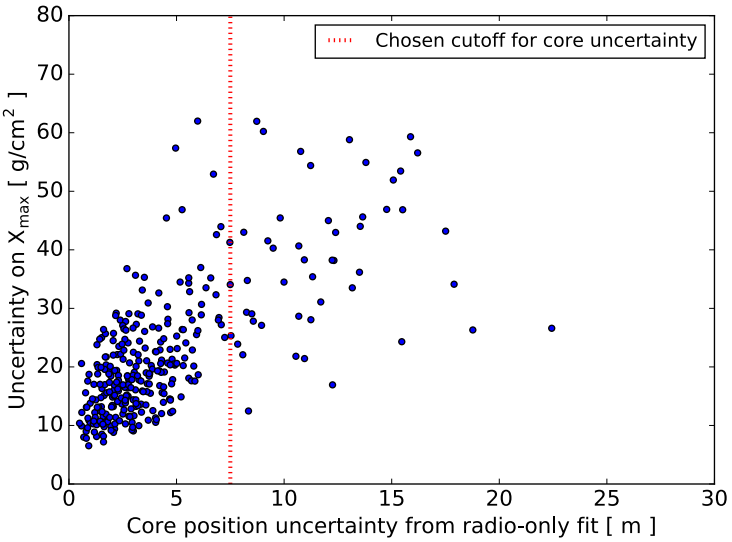


Figure 5.2: The uncertainty on $X_{\max}$ versus the uncertainty on the core position, per measured shower, after cuts on energy and fiducial selection criteria. The dotted line indicates the chosen cutoff.

The choice of hadronic interaction model also causes another, larger uncertainty in the composition analysis, as the average $X_{\max}$ for a given element varies by up to about $15\,\mathrm{g/cm^2}$ between models. This is treated separately by repeating the composition analysis with different models.

Residual uncertainties due to variations in the atmosphere, local weather etc. are about $2\,\mathrm{g/cm^2}$ from the 5-layer approximation of CORSIKA. An additional $4\,\mathrm{g/cm^2}$ is added in quadrature to the statistical uncertainties. A systematic uncertainty, or bias, in averages of $X_{\max}$ may arise from possible residual bias after applying the above selection criteria. We test this in Sect. 5.4.2, obtaining a value of $3.2\,\mathrm{g/cm^2}$ to be added as a systematic uncertainty on $X_{\max}$. Hence, a total systematic uncertainty on $X_{\max}$ of $7\,\mathrm{g/cm^2}$ follows. This is comparable to the systematic uncertainty on $X_{\max}$ in the measurements of Aab et al. (2014a) who find a value between 7 and $10\,\mathrm{g/cm^2}$ for primary energies above $10^{17.8}\,\mathrm{eV}$.

When performing the parabolic fit to the χ^2 values per simulation, as in the

Table 5.1: Systematic uncertainies in the $X_{\max}$ reconstruction

	Syst. uncertainty	Added stat. unc.
Choice of hadronic interaction model	$5\,\mathrm{g/cm^2}$	
Remaining atmospheric uncertainty	$\sim 1\,\mathrm{g/cm^2}$	$\sim 2\,\mathrm{g/cm^2}$
Five-layer atmosphere CORSIKA	$2\,\mathrm{g/cm^2}$	$4\,\mathrm{g/cm^2}$
Possible residual bias	$3.2\,\mathrm{g/cm^2}$	
Curve fit for χ^2 optimum	$\leq 1\,\mathrm{g/cm^2}$	
Total, added in quadrature	$7\,\mathrm{g/cm^2}$	

middle panel of Fig. 5.1, a systematic error of up to $5\,\mathrm{g/cm^2}$ may arise if the fit optimum is not contained in the dense region of simulations. This is removed by simulating extra showers around the optimum when needed. A Monte Carlo simulation shows no residual systematic error ($\leq 1\,\mathrm{g/cm^2}$) if the dense region is positioned asymmetrically around the optimum but does contain it.

The systematic uncertainty in the energy estimate from the LORA particle detectors was found to be 27 %, or 0.104 in $\lg(E)$ (Thoudam et al., 2014, 2016); by convention we write $\lg E \equiv \log_{10} E$.

5.3 Composition analysis

Having established the set of showers for the composition analysis, we perform statistical analysis on the measured data, being ($X_{\max}$, σ_X, $\lg E$, $\sigma_{\lg E}$) per shower. We make use of the probability density functions of $X_{\max}$ as a function of energy and atomic mass number A, as parametrized by De Domenico et al. (2013). The parametrizations follow the generalized Gumbel distribution, which is a function with 3 parameters, yielding a variable mean, spread, and tail-end asymmetry, respectively. The function has been fitted to a large sample of CONEX showers and has a precision within $2\,\mathrm{g/cm^2}$ for both average and standard deviation of $X_{\max}$, as well as a close fit to the distribution itself; the high-end tail, which drops exponentially, was shown to be well represented. It should be noted that CONEX is a faster but less accurate shower simulation method than CORSIKA. Average $X_{\max}$ values were found to deviate by 4 to $5\,\mathrm{g/cm^2}$. This is therefore treated as an additional

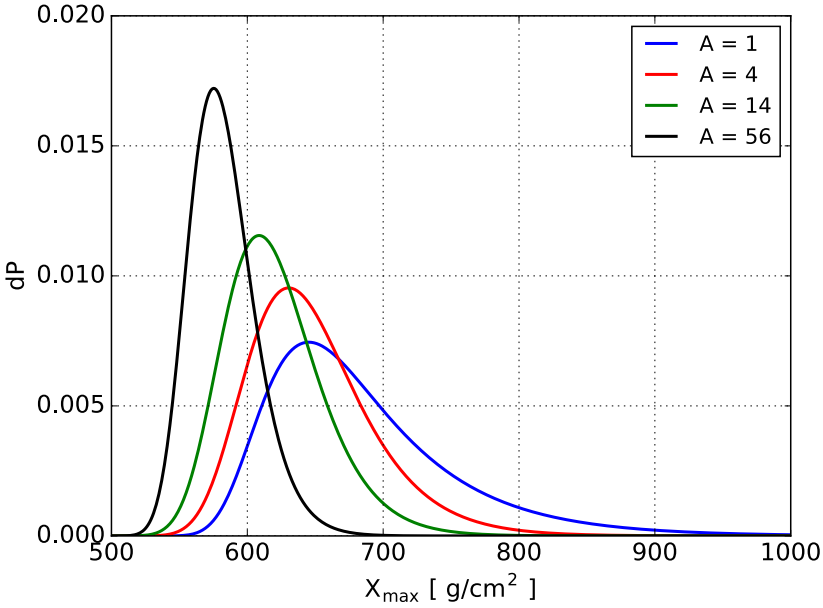


Figure 5.3: The probability density functions for the depth of shower maximum $X_{\max}$, for the elements H, He, N, and Fe, at energy 10^{17} eV and hadronic interaction model QGSJetII-04.

systematic uncertainty on $X_{\max}$, which for the composition analysis then amounts to $8\,\mathrm{g/cm^2}$.

Example curves are shown in Fig. 5.3 for $E = 10^{17}$ eV for protons, helium, nitrogen, and iron. The functions overlap substantially, limiting the extent to which the individual elements can be distinguished. This is the statistical challenge in performing a composition analysis on $X_{\max}$ data. The mean $X_{\max}$ shifts approximately proportionally to $\ln A$. Therefore, for a 4-component model of astrophysically relevant elements, a reasonable choice is to take p, He, C/N/O, and Fe. These are roughly equally spaced in $\ln A$; as C, N and O cannot be readily distinguished, either of them can be chosen as a proxy for all three. We choose nitrogen, as this is in between carbon and oxygen, and the best choice in terms of equal spacing in $\ln A$.

5.3.1 Unbinned maximum likelihood analysis

The considerable overlap of the $X_{\max}$ distributions calls for a statistical method that has optimal distinguishing power. An unbinned maximum likelihood analysis is optimal in the large-N limit (e.g. Bohm and Zech, 2010), and treats each shower separately instead of relying on histograms and/or binning in energy. This is especially suitable when the dataset is relatively small and e.g. a narrow binning in shower energies is inappropriate.

A downside of this method is that it does not deliver a goodness-of-fit estimate, i.e. it will select the best-fitting mixed composition model, but does not directly inform how well this model fits the data. Therefore, a complementary goodness-of-fit test is done separately as described in Sect. 5.3.3.

Given a measured shower with parameters ($X_{\max}$, σ_X, $\lg E$, $\sigma_{\lg E}$), its likelihood function for a given mixed mass composition is

$$L\left(\{\alpha_i\};\ X_{\max},\, E\right) = \sum_i \alpha_i\, f_i(X_{\max},\, E) * \mathcal{N}_X(X_{\max},\, \sigma_X^2) * \mathcal{N}_{\lg E}(\lg E,\, \sigma_{\lg E}^2), \quad (5.2)$$

where the model elements are indexed by i, their mix fractions are $\{\alpha_i\}$, and $f_i(X_{\max}, E)$ is the probability density function for element i over $X_{\max}$, at energy E, as taken from De Domenico et al. (2013). Because of the measurement uncertainties in $X_{\max}$ and energy, the resulting function is convolved (the $*$ symbol in Eq. 5.2) with a corresponding Gaussian in each parameter.

The likelihood function for the set of showers is the product of L for each shower. As is customary, we use the log-likelihood, being the sum of $\ln L$ for all showers. The maximum likelihood method is defined as finding the mixed composition $\{\alpha_i\}$ which maximizes the total $\ln L$:

$$\max_{\{\alpha_i\}} \sum_{k=1}^{N} \ln L(\{\alpha_i\};\ X_{\max,\mathrm{k}},\, E_k), \quad (5.3)$$

where the index k denotes the sum over all showers.

We find the best-fitting composition using a 2 percent-point grid search over the 3 independent parameters (the four mix fractions must sum up to exactly 1), followed by a Nelder-Mead simplex optimization for sub-grid precision. Omitting the grid search may produce a local optimum away from the global optimum, as we have found.

As evaluating the likelihood function Eq. 5.2 is relatively expensive due to the double convolution, we note that the likelihood for a mixed composition is a linear combination of the likelihood functions for each element separately. Therefore, these need to be calculated only once for each measured shower, to be tabulated for quick reference later. Therefore, our grid search takes only about 2 seconds for 196 measured showers (Python implementation without further optimizations), and scales linearly. The same approach is thus practical for large data sets as well, and also for a 5-component model or, perhaps with a parallel implementation, 6 components.

5.3.2 Uncertainty analysis using the likelihood ratio test

Having found the maximum likelihood estimate for the composition, we need to find the uncertainties or confidence intervals for each of the elements. It can also be done for groups of elements such as protons plus helium combined (light versus heavy particles). The likelihood ratio test is a reliable way to obtain these in a frequentist setting. Alternatively, one could opt for a Bayesian analysis using the likelihood function results from the grid search.

Denoting the likelihood of the best-fitting composition as $L(\{\hat{\alpha}\})$, we fix one of the element fractions, e.g. the proton fraction, scanning over the range from 0 to 1. We then find the maximum likelihood composition given the fixed proton fraction, $L(\alpha_p, \{\hat{\alpha}_i\})$, again optimizing over the free parameters indexed by i. This gives the test statistic D:

$$D = 2 \ln \left(\frac{L(\{\hat{\alpha}\})}{L(\alpha_p, \{\hat{\alpha}_i\})} \right). \tag{5.4}$$

This is nonnegative by definition, and by Wilks' theorem (Bohm and Zech, 2010), in the large-N limit it follows a chi-squared distribution with 1 degree of freedom, when fixing 1 parameter. If this theorem is well applicable given the size of our dataset, the confidence intervals at significance level $1 - p$ follow directly from the critical values of the chi-squared distribution. For confidence levels of 68, 95, and 99%, these are 1.00, 3.84, and 6.64, respectively.

We have tested the validity of Wilks' theorem for our case, where N is finite and the parameters are limited to the range 0 to 1. We have produced $\sim 10^4$ simulated datasets given the best-fit composition $\{\hat{\alpha}_i\}$. Each of them was analyzed

to produce its own best-fit composition and corresponding likelihood value. Keeping one parameter fixed, e.g. the best-fit proton fraction from the data $\hat{\alpha_p}$, we produce the best-fit composition given $\hat{\alpha_p}$, and compute D from Eq. 5.4. This is repeated for the other elements given $\hat{\alpha_k}$.

We have found that for our dataset, D closely follows the chi-squared distribution whenever $\hat{\alpha_k}$ is significantly away from the boundaries at 0 and 1, with a difference in 99 % critical value only on the order of 0.2. Already at a fraction of 0.05, such as fitted for iron (see Sect. 5.5.1), the critical value is only 0.4 lower than expected. However, when the best-fit value is zero, it deviates (becomes lower), and we take the critical values from these simulations.

Confidence intervals for two elements simultaneously, such as used in the contour plot Fig. 5.10 in Sect. 5.5 are computed analogously, fixing two parameters instead of one, and noting that the test statistic D then follows a $\chi^2(2)$-distribution.

5.3.3 Goodness of fit

The analysis as presented so far, does not establish whether the best-fitting composition model is a good, poor or bad fit to the measured $X_{\max}$ distribution. One cannot use the value of the (log)likelihood function for this; for instance, a trivial 'dataset' with N identical $X_{\max}$ values could produce either a low, moderate or high likelihood value, while none of the composition models would be a good fit.

Therefore, we perform a separate goodness-of-fit test. The Kolmogorov-Smirnov test is a well-known method for this (Bohm and Zech, 2010), comparing the cumulative distribution function (cdf) of the best-fit model to the empirical cumulative distribution of the data. Using cumulative distributions is useful, as unlike a histogram, the empirical distribution for a dataset is a uniquely defined and complete description of the data. The empirical distribution is defined as

$$E(X) = \frac{1}{N} \sum_{i=1}^{N} \theta\,(X - X_i), \tag{5.5}$$

with $\theta(X)$ the unit step function. Effectively it counts the number of data points below X.

The best-fit model $X_{\max}$ distribution is taken as the sum of the distributions as taken from the likelihood function for each of the measured showers:

$$f_{\text{sum}}(X_{\max}) = \frac{1}{N}\sum_{k=1}^{N}\sum_{i}\alpha_i\, f_i(X_{\max},\, E_k) * \mathcal{N}_X(X_{\max},\, \sigma^2_{X,k}) * \mathcal{N}_{\lg E}(\lg E_k,\, \sigma^2_{\lg E,k}), \tag{5.6}$$

summing over all showers (index k) and over the elements in the composition model (index i). The measured showers vary in energy, energy uncertainty $\sigma_{\lg E}$, and in $X_{\max}$-uncertainty σ_X, hence the index k in E_k, $\sigma_{X,k}$, and $\sigma_{\lg E,k}$ in this equation. Consequently, random sampling from the sum distribution includes a resampling of the values of E, σ_X, and $\sigma_{\lg E}$ in the dataset, such as is done e.g. in a non-parametric bootstrap.

The Kolmogorov-Smirnov test statistic is defined as the maximum difference between the model's cdf $F(X)$ and the empirical distribution $E(X)$:

$$K = \sup_X |F(X) - E(X)|\,. \tag{5.7}$$

It is in fact a distance measure between F(X) and E(X), using the L_∞-norm.

We test the null hypothesis that the data is a random drawing from the best-fitting composition model $F(X)$. To this end, we use inverse-cdf sampling to simulate 10^5 size-N $X_{\max}$ datasets drawn from $F(X)$. For each set, we compute K from Eq. 5.7. The sorted set of K-values yields the critical value of K for the desired significance level. Analogously, the p-value corresponding to K for the measured dataset follows from the percentile level of K_{data} in the set of simulated K-values.

The set of simulated empirical distributions for which K is below its 95-percentile value is used to visualize a band around $F(X)$ in which the dataset's $E(X)$ should fall to have $p > 0.05$ for the null hypothesis.

5.4 Results: the measured $X_{\max}$ distribution

In the following sections, we present the results regarding statistics on $X_{\max}$, such as the estimate of the mean and standard deviation of the $X_{\max}$ distribution. After this, the implications for the cosmic-ray composition in our energy range are given, based on the statistical analysis presented in Sect. 5.3.

The results are based on a dataset of $N = 196$ cosmic rays with energies between 10^{17} and 10^{18} eV, which pass all selection criteria for a bias-free sample with accurately reconstructable showers, as explained in Sect. 5.2.4. The uncertainty on the $X_{\max}$ measurement per shower is on average $19\,\mathrm{g/cm^2}$. The average fit quality of the best-fitting simulation to the measured LOFAR data is $\chi^2/\mathrm{dof} = 1.20$, indicating a good fit.

We have excluded a time window from September 2015 through February 2017, when four LORA particle detectors malfunctioned for significant time intervals. The exact cause, and the consequences for the fiducial selection criteria are still under investigation, hence the safest approach is to exclude these data.

5.4.1 Mean and standard deviation of $X_{\max}$ as a function of primary energy

We have divided the dataset into four energy bins, and computed the mean and standard deviation in each bin. The sample averages are shown in Fig. 5.4. The given uncertainty is the uncertainty on the mean of the $X_{\max}$ distribution, i.e. $\sigma/\sqrt{N}$, with sample standard deviation σ. For comparison, results are included from Pierre Auger (Bellido et al., 2017), HiRes (Sokolsky, 2011), Tunka (Prosin et al., 2015), and Yakutsk (Knurenko et al., 2015).

The result in the first bin is higher by $12\,\mathrm{g/cm^2}$, and in the second bin is lower by $11\,\mathrm{g/cm^2}$, compared to the results presented in Buitink et al. (2016). The differences can be explained through statistical fluctuations, and from the revised treatment of systematic effects including the atmosphere and the radio-derived energy scale.

The average $X_{\max}$ agrees well with the other experiments such as Tunka and Yakutsk, and with HiRes/Mia up to $\lg E \sim 17.7$. However, the results from the Pierre Auger Observatory, which is the largest experiment, are significantly higher. Their statistical uncertainty is smaller than the plotted symbols, arising from a very high number of showers (1000 to 2600) per bin. Systematic uncertainties on $X_{\max}$ in this energy range are about $11\,\mathrm{g/cm^2}$ for Auger (Bellido et al., 2017), and about $7\,\mathrm{g/cm^2}$ for LOFAR. Additionally, there is a systematic uncertainty in energy, which for LOFAR is about 0.10 in $\lg E$. As explained in Sect. 5.5.2, such a shift in energy would lead to a shift in $\langle X_{\max}\rangle$ of about $6\,\mathrm{g/cm^2}$ due to the natural trend of $\langle X_{\max}\rangle$ with energy (i.e. the elongation rate).

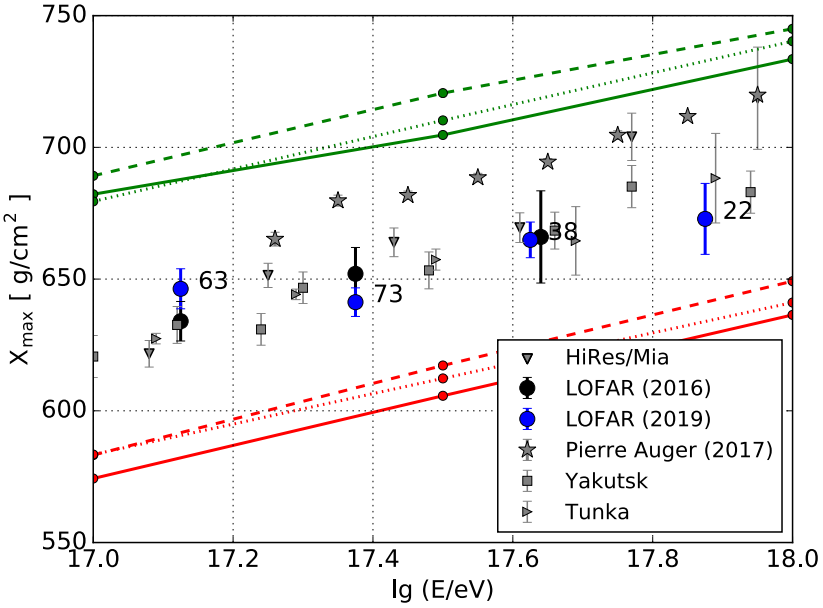


Figure 5.4: The average depth of shower maximum $X_{\max}$, as a function of primary particle energy. The annotated numbers indicate the number of showers in each bin, and the error margins indicate the uncertainty on the mean of the $X_{\max}$ distribution. The upper lines indicate the mean values expected for protons, from simulations with QGSJetII-04 (solid), EPOS-LHC (dashed) and Sibyll-2.1 (dotted). The lower lines show the mean predicted values for iron nuclei. For comparison, results from Pierre Auger, Yakutsk, Tunka, and HiRes/Mia are included.

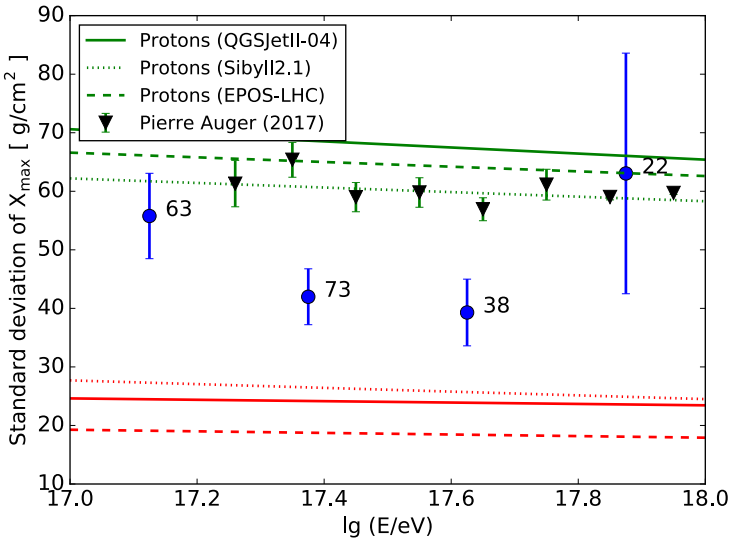


Figure 5.5: The standard deviation of $X_{\max}$ as a function of primary particle energy. The margins indicate the uncertainty on the standard deviation of the $X_{\max}$ distribution. Results from the Pierre Auger Observatory are shown, together with the values from simulations of protons and iron nuclei (high and low lines, respectively).

Therefore, most of the discrepancy is explainable within systematic uncertainties. However, there is a notable difference in methodology to measure $X_{\max}$, direct fluorescence detection versus radio detection with Corsika/CoREAS simulations. The measured differences in average $X_{\max}$ indicate that a detailed comparison between experiments, of the measurements and their systematic effects, would be recommended in future research.

In Fig. 5.5, we show the standard deviation in each bin, along with its uncertainty. To calculate these, as an estimator $\hat{\sigma}$ of the underlying $X_{\max}$ distribution's standard deviation, we subtract the variance caused by the $X_{\max}$ uncertainty per measured shower:

$$\hat{\sigma} = \sqrt{\sigma^2 - \frac{1}{N_j}\sum_{i=1}^{N_j} u_i^2}, \tag{5.8}$$

with σ the sample standard deviation, u_i the $X_{\max}$ uncertainty on each shower, and N_j is the number of showers in energy bin j.

The uncertainty on the standard deviation of the distribution is estimated using (nonparametric) bootstrap resampling. As the figure shows, especially in the highest bin the uncertainty is large, owing to the relatively low number of showers per bin. The results do give an indication, as do the averages, of values in the middle between those for a pure proton and iron distribution, respectively.

The results are lower than those from the Pierre Auger Observatory, as follows especially from the middle two energy bins. This implies a narrower measured distribution. Both a lower average $X_{\max}$ and a lower variance are in line with a somewhat heavier composition, although a lower variance could also arise from a more pure mixture, e.g. by having one element that dominates in the composition.

5.4.2 Tests for residual bias

The average $X_{\max}$ of the measured showers is a given quantity that does not depend on shower parameters such as the zenith angle. This follows from the fact that the cosmic-ray composition in our energy range is independent of time and incoming direction, as far as is known from experiments.

This allows to perform a test of our sample for a residual bias in $X_{\max}$ due to variations in these parameters. Our sample of 196 showers has an average $X_{\max}$ of $651 \pm 4\,\mathrm{g/cm^2}$. As discussed in Sect. 5.2.4, a biased sample would readily show a dependence of the average $X_{\max}$ on zenith angle. However, the average $X_{\max}$ also depends on the energy; its expected value is to good approximation linear in $\lg E$ over the range from 10^{17} to $10^{18}\,\mathrm{eV}$. From the parametrization using Gumbel distributions, as discussed in Sect. 5.3, and for the QGSJetII-04 hadronic interaction model, we find, for a factor 10 increase in energy, an average rise in $X_{\max}$ (also referred to as elongation rate) of 54.3, 55.6 and $59.2\,\mathrm{g/cm^2}$, for protons, nitrogen, and iron nuclei, respectively. This is in good agreement with the elongation rate of $58\,\mathrm{g/cm^2}$ predicted e.g. by the Heitler-Matthews model (Matthews, 2005).

We have therefore evaluated the average log-energy versus zenith angle, in six bins, as shown in the left panel in Fig. 5.6. It shows a trend towards higher energy with increasing inclination of the showers, rising by about 0.23 from 10 to 45 degrees of inclination. A possible residual bias in $X_{\max}$, corrected for the influence of varying

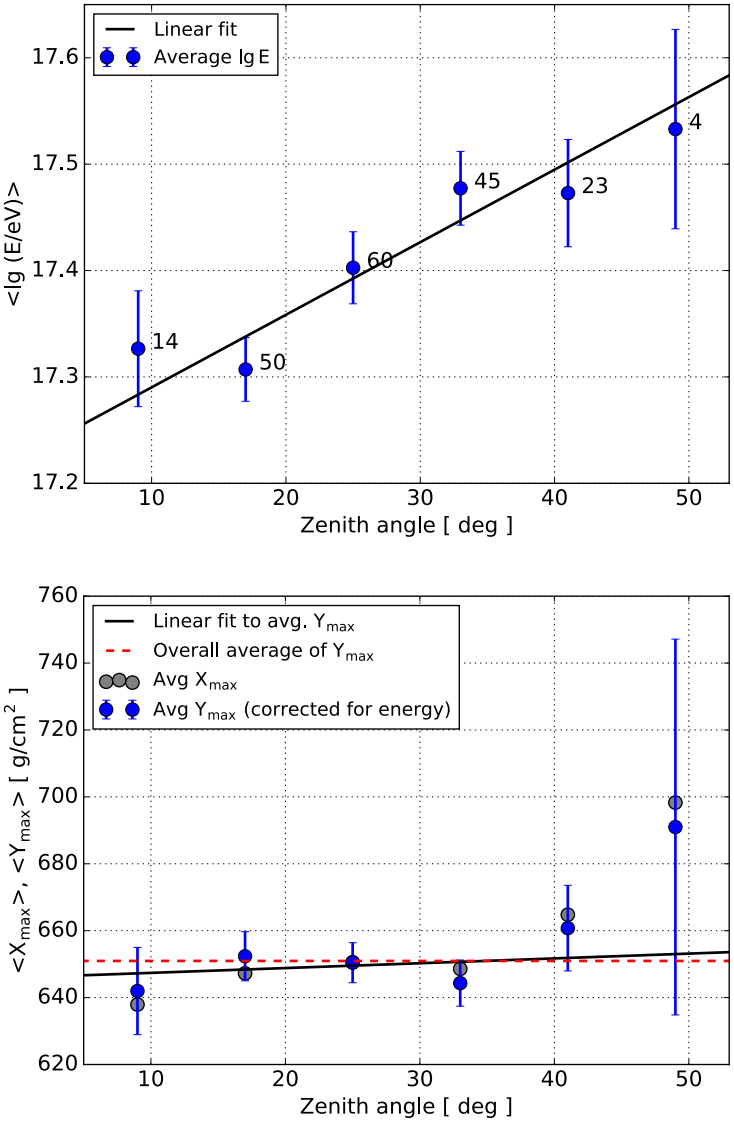


Figure 5.6: Left panel: The average lg E as a function of zenith angle, together with a linear fit. Right panel: Average Y (from Eq. 5.9) versus zenith angle, together with a constant and linear fit.

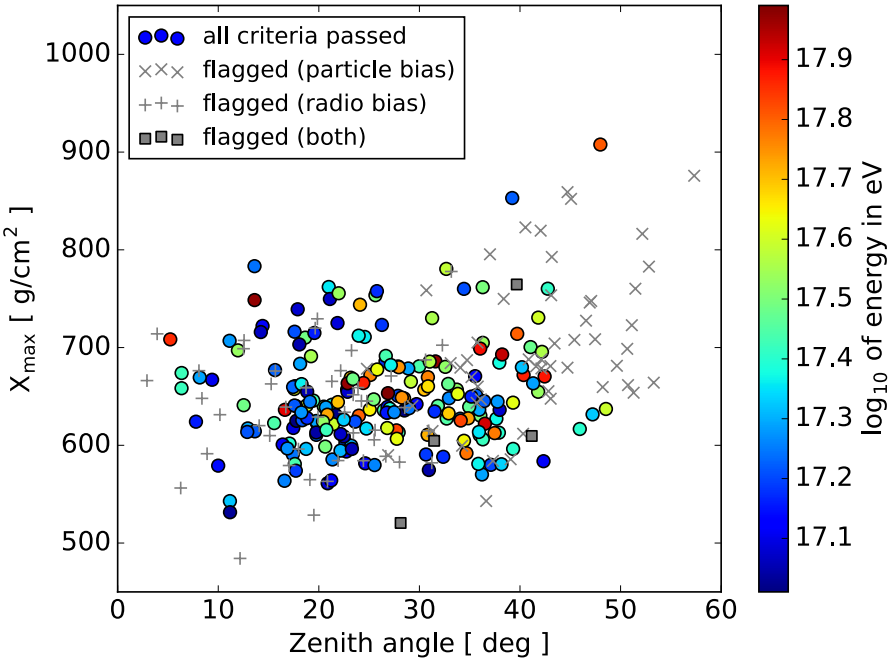


Figure 5.7: Scatter plot of $X_{\max}$ versus zenith angle, for all 298 showers with core reconstruction precision better than 7.5 m. Colored circles represent the 196 showers passing all criteria, with the color denoting their energy. The showers flagged by the particle and radio bias tests are also shown.

energy, is evaluated by introducing a parameter Y for each shower, as

$$Y = X_{\max} + 55\,(\lg(E) - 17.4)\ \mathrm{g/cm^2}, \tag{5.9}$$

where 17.4 is approximately the average value of log-energy in our sample.

The results are shown in the right panel of Fig. 5.6, together with a linear fit. The uncertainty margins are once again given by the standard error of the mean. A constant fit of $Y = 650$ as well as a linear fit are shown.

The linear fit has a slope parameter of 0.14 ± 0.44. Hence, at our level of statistics, the slope is statistically compatible with zero, and no residual bias is evident. The

high value of about 690 for the rightmost bin appears suggestive, but as it contains only 4 showers and has a correspondingly large uncertainty, it is not significant.

The constant fit has an uncertainty of $3.2\,\mathrm{g/cm^2}$. A bias at this level cannot be ruled out, hence this is added as a contribution to the systematic uncertainty on X_{max}.

We also show a complete scatter plot of the set of showers versus zenith angle, in Fig. 5.7. This plot shows the effect of the bias tests for the radio and particle detectors and the corresponding fiducial cuts (Sect. 5.2.4.1 and 5.2.4.2). As expected, the particle bias test flags most events at high inclination and high X_{max}, above 45 degrees. The radio bias test flags mostly the opposite region, low X_{max} and low inclination.

Consequently, we see only few showers passing the tests at $\theta < 10°$, and there are only 4 in the highest zenith angle bin above 46°. The plot makes clear that the fiducial cuts from Sect. 5.2.4 are necessary, as there would have been a strong zenith angle dependence, and thus a biased X_{max}-sample, had it been omitted.

5.5 Composition results

We have applied the statistical analysis in Sect. 5.3 to the set of 196 showers passing all criteria. The details of the analysis are presented for the QGSJetII-04 hadronic interaction model in the next section. After discussing systematic uncertainties, the results for this model as well as EPOS-LHC (Pierog et al., 2015) and Sibyll-2.1 (Ahn et al., 2009) are discussed in Sect. 5.5.3. The results for the three hadronic interaction models are summarized in Fig. 5.11.

5.5.1 Statistics for the QGSJetII-04 hadronic interaction model

The maximum likelihood estimate was found to be 17 % protons, no helium, 78 % nitrogen and 5 % iron. A histogram of X_{max} is shown in Fig. 5.8 (top), for the full energy range of 10^{17} to 10^{18} eV. The red (solid) curve is the best-fitting distribution, found using the maximum likelihood method and Eq. 5.6. The distributions for the three elements that make up the best-fitting distribution are also shown, scaled by their respective mix fractions.

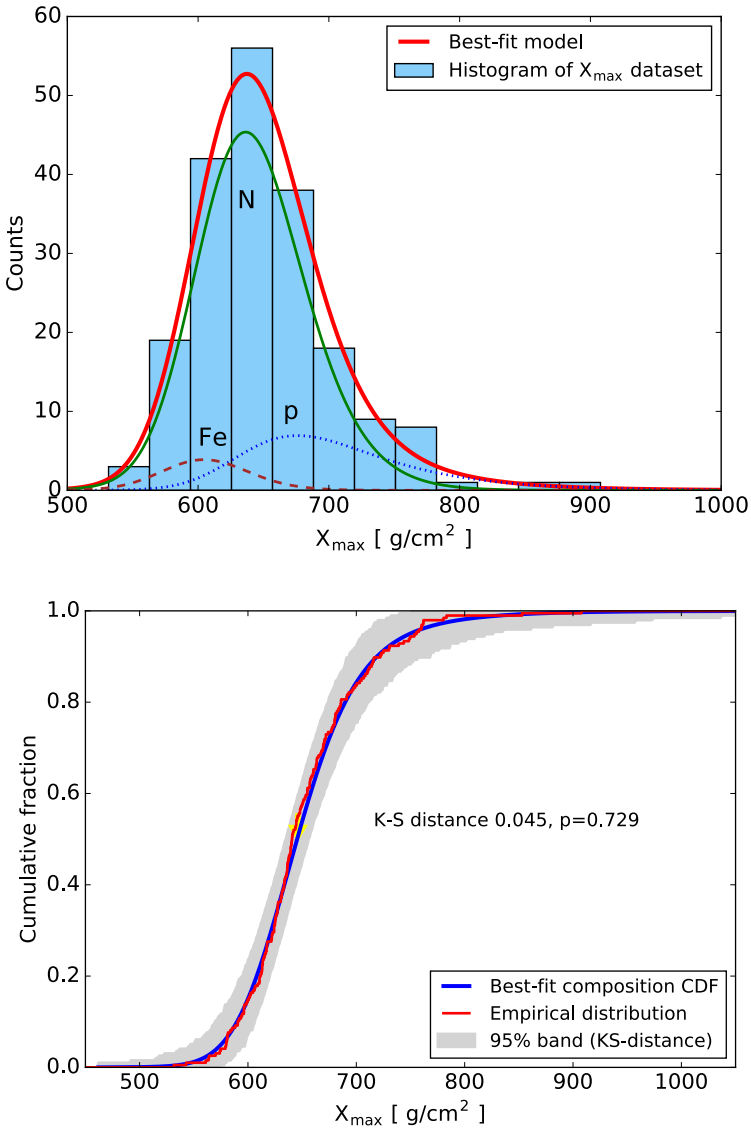


Figure 5.8: Top: A histogram of $X_{\max}$, together with the best-fitting distribution from Eq. 5.6. The helium fraction was zero at best fit. Bottom: The cumulative distribution corresponding to the best-fit composition model, together with the empirical distribution from our dataset. The grey band is the envelope of all simulated empirical distributions that have a K-S distance to the CDF below its 95-percentile level.

We have tested the goodness-of-fit of the best-fitting model to our dataset, using the cumulative and empirical distributions and the Kolmogorov-Smirnov test (Sect. 5.3.3). As shown in Fig. 5.8 (bottom), the model is a good fit to the data ($p = 0.729$). Switching the hadronic interaction model to EPOS-LHC or Sibyll-2.1 produces about equally good fits to the data, at $p = 0.81$ and $p = 0.82$ respectively. Hence, all three models fit the data well, at their respective best-fitting composition. It is of course possible that this would change with a larger dataset.

Fig. 5.9 shows the results of the uncertainty analysis based on the likelihood ratio test, for the four elements in our composition model. We have plotted the test statistic D from Eq. 5.4 with a minus sign, to find the confidence intervals where the curves are above the dotted line corresponding to the desired confidence level.

This yields the confidence intervals for the element fractions, i.e. the statistical uncertainties. For example for protons, the 95 % confidence interval runs from 0 to 23 %; the 99 % confidence interval runs up to 30 %. Similarly, an upper bound for the iron fraction follows at 14 or 24 %, at 95 or 99 % confidence level, respectively. The one-sigma (68 %) confidence interval is found by considering a critical level of 1.0 for the D-statistic. We observe that protons and helium are to a significant degree interchangeable in the statistical analysis of our model, given our dataset. This is readily seen in the contour plot in Fig. 5.10, showing the D-statistic versus proton and helium fractions. The contours show the allowed regions with confidence levels one-sigma (i.e. 68 %), 95 %, and 99 %, respectively, with their D-values being the exceeding probabilities of a χ^2-distribution with two degrees of freedom. Within the one-sigma region, one can exchange helium for protons in a ratio of about 3 to 1.

5.5.2 Accounting for systematic uncertainties

The systematic uncertainty in $X_{\max}$ amounts to $\pm 8.1\,\mathrm{g/cm^2}$ (before roundoff), including the contribution from the CONEX-based parametrizations. The energy uncertainty of 27 %, or 0.104 in $\lg E$ has, to first order, the effect of an overall shift of $X_{\max}$ in the $X_{\max}$-distributions (see Eq. 5.9), of $5.7\,\mathrm{g/cm^2}$. By adding both uncertainties in quadrature, we obtain a systematic uncertainty of $9.9\,\mathrm{g/cm^2}$.

Evaluating the composition results for $X_{\max} \pm 9.9\,\mathrm{g/cm^2}$ for all showers, we obtain limits for the best fit, as well as the (expanded) confidence intervals that

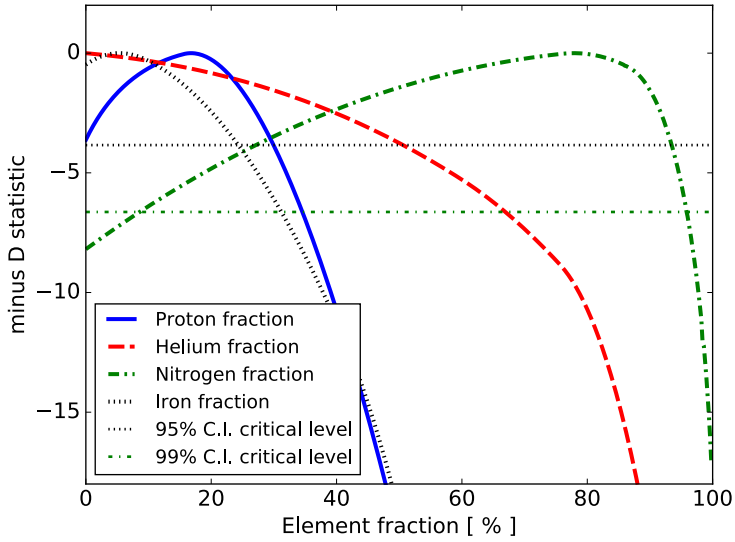


Figure 5.9: Applying the likelihood ratio test, giving the D-statistic found from Eq. 5.4. Confidence intervals are defined as the region where the curves are above the dotted lines. For the 68 % confidence interval, i.e. one sigma, the critical level is -1.0 in this plot.

arise for a systematic shift in this range. Noteworthy is that when the average $X_{\max}$ is shifted upward, e.g. the helium fraction is fitted much higher at the expense of the nitrogen fraction. Helium is then favored over nitrogen in the fit, due to its higher expected $X_{\max}$.

5.5.3 Results for three hadronic interaction models

The results for the hadronic interaction models QGSJetII-04, EPOS-LHC, and Sibyll-2.1 are plotted in Fig. 5.11. The results for QGSJetII-04 and Sibyll-2.1 are almost the same. The analysis provides an upper bound to the fraction of protons of 40 % at 95 % confidence level, or 43 % at 99 % confidence level, and within systematic uncertainties. This result holds for all three hadronic interaction models we considered.

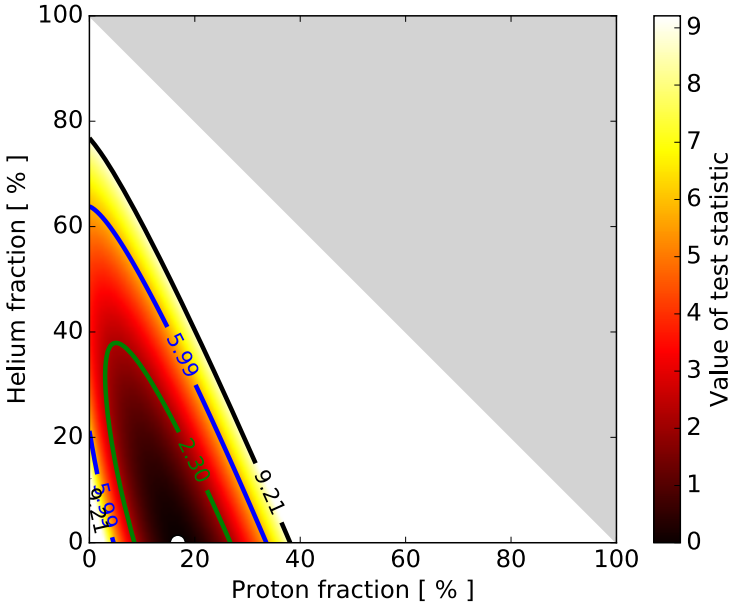


Figure 5.10: Contour plot of the proton and helium fraction, giving the regions consistent with a one-sigma, 95 % and 99 % confidence level, respectively.

For EPOS-LHC, it is seen that the fit favors a more heavy composition, with more iron instead of nitrogen, and with a significant upper bound on the helium fraction. Otherwise, the results are very similar across these three hadronic interaction models; it is clear that the intermediate-mass elements in the C/N/O range (possibly stretching to somewhat higher mass numbers as well) are dominant. Apart from a crossover from C/N/O to iron, the choice of hadronic interaction model has only a limited effect on the best-fit results.

It is clear that the composition cannot be described as a two-component mixture of protons and iron; however, a two-component mixture of protons and nitrogen would work for QGSJetII-04 and Sibyll-2.1.

For this dataset, helium and nitrogen are not fully resolved in the composition model. The fitted values for the helium and nitrogen fraction are highly anti-

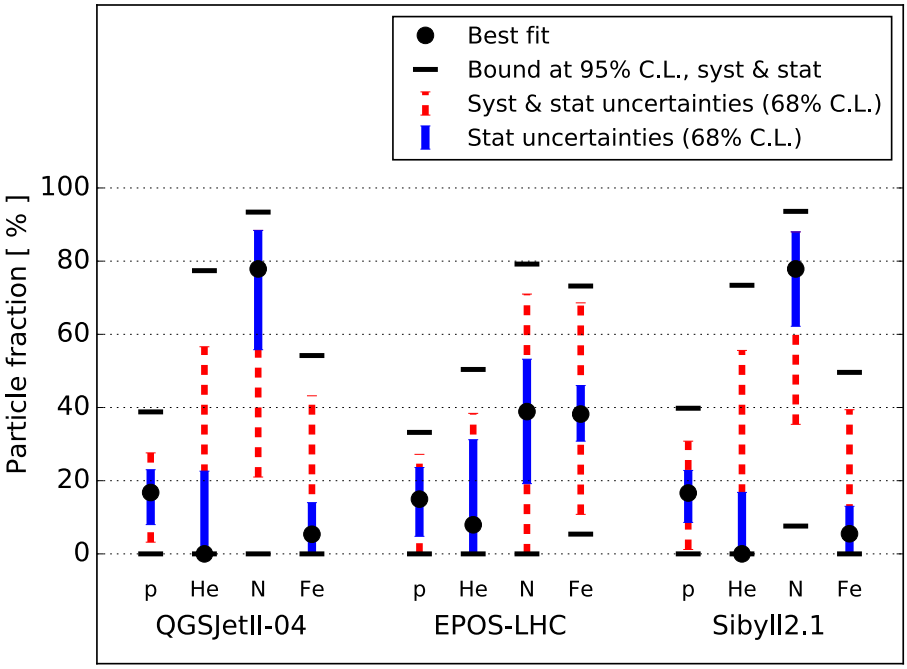


Figure 5.11: Composition results from our dataset, assuming each of the three hadronic interaction models shown at the bottom. The best fit is shown along with statistical and systematic uncertainties, and bounds for the 95 % confidence interval including systematic uncertainties.

correlated, which follows from the requirement that all mix fractions sum up to 1. Helium and nitrogen are a factor 3.5 apart in atomic mass number, whereas the other consecutive elements are a factor 4 apart; constant factors here correspond to a constant shift in $\ln A$, and the means of the $X_{\max}$-distributions differ by an amount proportional to $\ln A$. Moreover, helium and nitrogen have two 'neighboring' elements in the composition model, unlike hydrogen and iron. This increases sensitivity to a systematic shift (up or down) in $X_{\max}$.

Generally, the statistical margins indicate that the analysis would improve with more data. This is no surprise, at our modest level of statistics. However, we see that the systematic uncertainties are also important already at our level of statistics, even though they are comparable to other experiments such as the Pierre Auger Observatory. This is especially evident in the fitted iron fraction, which is well bounded by statistics, but has substantially expanded margins when systematic uncertainties are included. Also, looking once more at the substantially overlapping $X_{\max}$-distributions in Fig. 5.3, it is clear that achieving lower systematic offsets in $X_{\max}$ is still important, to improve the resolution of the composition analysis.

5.5.4 Variations with energy: composition analysis in two energy bins

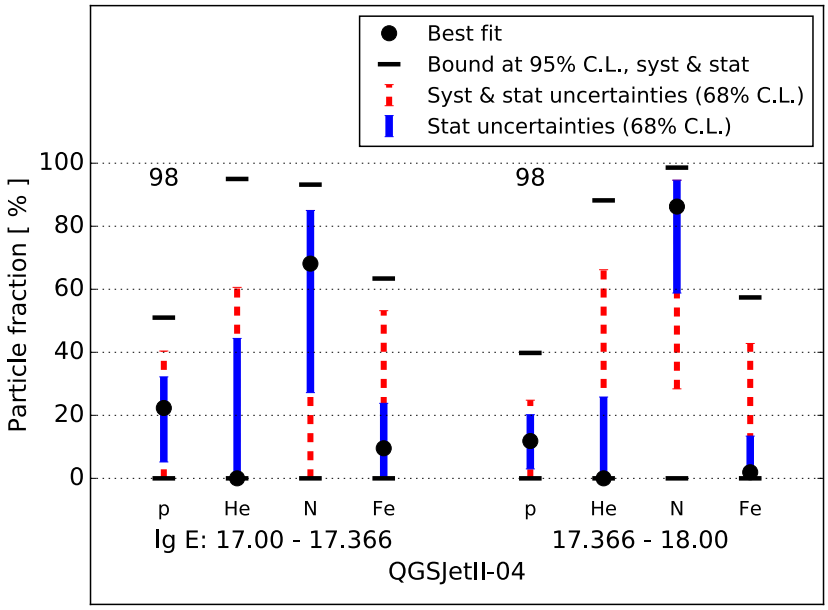
For evaluating possible variations of the composition with primary energy, a first step is to divide the sample into two bins with an equal number of showers. As our dataset samples the energy interval 10^{17} to 10^{18} eV non-uniformly due to the natural fall-off in the cosmic ray spectrum, the two bins will not be equally wide in (log-)energy. However, having an equal number of showers is the best choice for preserving the level of statistics.

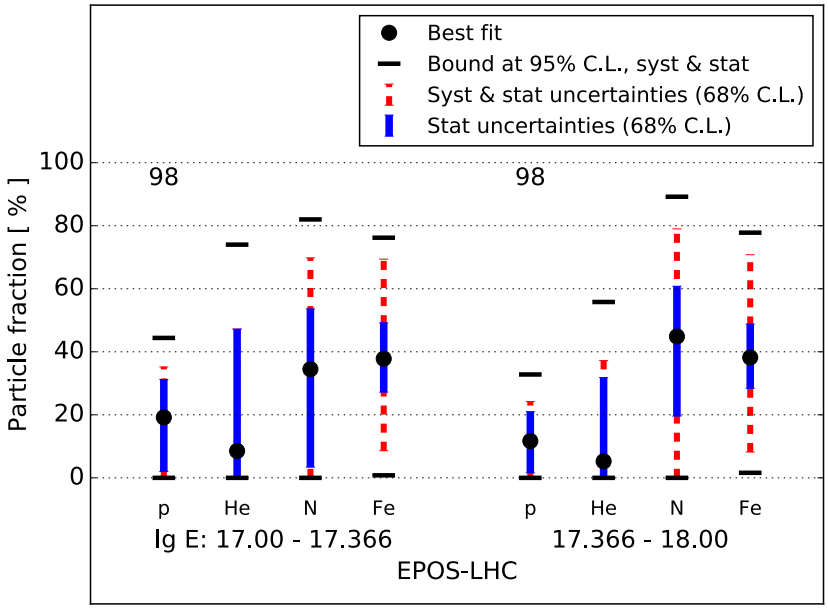
The composition results for the three hadronic interaction models are shown in Fig. 5.11. Again, in both energy bins, a description of the composition requires the intermediate-mass elements such as C/N/O.

For all three models, and especially for EPOS-LHC, the results in the two bins are very similar, with differences much smaller than statistical margins. At best fit, protons and/or helium fractions are slightly lower in the high-energy bin, while the nitrogen fraction increases somewhat.

Hence, our dataset shows no evidence for a change in composition with energy,

either from the four-element analysis or from the average X_{max} as shown in Fig. 5.4. Whether the composition is actually constant with energy can only be settled with a larger data sample.


Best fit
Bound at 95% C.L., syst & stat
Syst & stat uncertainties (68% C.L.)
Stat uncertainties (68% C.L.)
Particle fraction [%]
100
80
60
40
20
0
98
98
p
He
N
Fe
p
He
N
Fe
lg E: 17.00 - 17.366
17.366 - 18.00
QGSJetII-04


Best fit
Bound at 95% C.L., syst & stat
Syst & stat uncertainties (68% C.L.)
Stat uncertainties (68% C.L.)
Particle fraction [%]
100
80
60
40
20
0
98
98
p
He
N
Fe
p
He
N
Fe
lg E: 17.00 - 17.366
17.366 - 18.00
EPOS-LHC

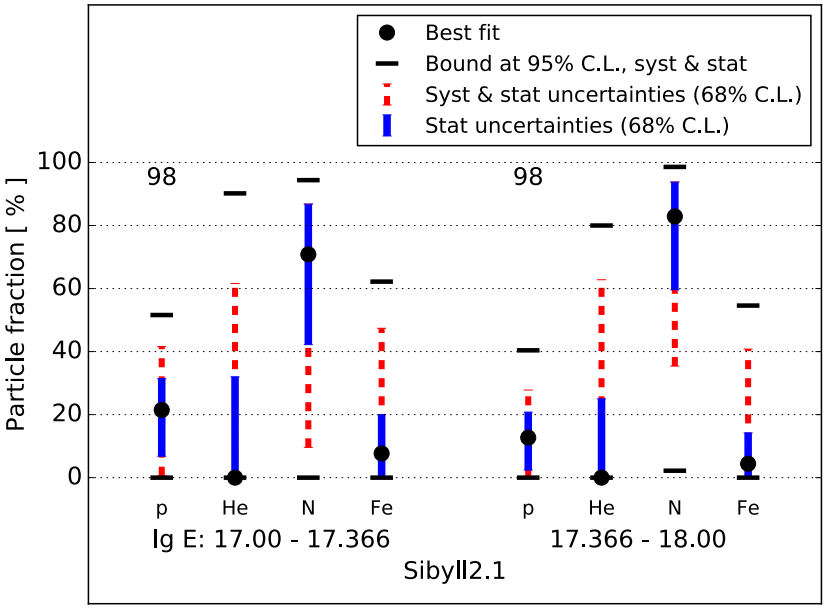


Figure 5.11: Composition results for two energy bins, below and above the median energy. The number of showers per bin is given near the top of each figure; the hadronic interaction model is shown at the bottom.

5.6 Summary

We have presented a method for estimating the composition of cosmic rays in the energy range 10^{17} to 10^{18} eV, based on LOFAR radio measurements of the depth of the shower maximum $X_{\max}$ of air showers. Using an ensemble of CORSIKA / CoREAS simulations per air shower, we infer its $X_{\max}$ at a resolution of $19\,\mathrm{g/cm^2}$ on average. The simulations are a close fit to the data, at $\chi^2/\mathrm{dof} = 1.20$ on average, for the best-fitting simulated shower. The systematic uncertainty on the $X_{\max}$-measurement amounts to $7\,\mathrm{g/cm^2}$, which is about the same as the uncertainty on measurements at the Pierre Auger Observatory above $10^{17.8}$ eV.

To obtain an unbiased sample suitable for composition studies, three selection criteria were applied. Requiring a shower core precision better than 7.5 m gives a sufficient general reconstruction quality cut. We further require that each simulated shower in the CORSIKA / CoREAS ensemble must be able to trigger the LORA particle detector array, and also pass the detection and quality criteria of the LOFAR radio analysis pipeline. This procedure leaves a sample of 196 showers for the composition analysis.

We have used an unbinned maximum likelihood method to obtain the best-fitting composition. This relies on high-precision parametrizations of the $X_{\max}$-distributions of the elements, as produced by De Domenico et al. (2013). As they use CONEX showers, which are an approximation to fully simulated CORSIKA showers, an additional systematic uncertainty on $X_{\max}$ of $5\,\mathrm{g/cm^2}$ is added in quadrature. A systematic energy uncertainty of 27 %, together with the systematic uncertainty on $X_{\max}$, translates into a total systematic uncertainty on $X_{\max}$ of $9.9\,\mathrm{g/cm^2}$ for the composition analysis.

A likelihood ratio test was used to establish the confidence intervals for the element fractions, after a separate test showed its validity in our case, with parameters bounded to the interval $[0, 1]$. As the unbinned analysis does not provide a goodness-of-fit estimate of the model distribution to the data, we have performed a Kolmogorov-Smirnov test. It shows a good fit to the data, at $p = 0.73$ for QGSJetII-04, $p = 0.81$ for EPOS-LHC, and $p = 0.82$ for Sibyll-2.1.

We have used a four-component model of elements, about equally spaced in $\ln A$. The best-fitting composition for our dataset is 17 % protons, no helium, 78 %

nitrogen and 5 % iron, assuming the QGSJetII-04 hadronic interaction model. The EPOS-LHC model tends towards a heavier composition, with an iron fraction of 38 % as a best fit. The Sibyll-2.1 model produces almost the same results as QGSJetII-04. Overall, we see only small differences in composition results between these three important hadronic interaction models, apart from some shift between nitrogen and iron.

The strongest composition result in terms of statistical significance is an upper bound on the proton fraction of 43 %, at 99 % confidence level, including systematic uncertainties, and including three different hadronic interaction models. Hence, the majority of cosmic rays at this energy must be heavier particles. It also follows that the cosmic-ray composition cannot be described by a two-component mixture of protons and iron; at least half of the particles are nuclei from the intermediate-mass elements, ranging from helium to (somewhat beyond) oxygen.

We have divided the sample into two bins, for energies above and below the median. The composition results for the two bins are quite similar, with only small differences at best fit. The differences between the two bins are not significant; also the observed average $X_{\max}$ values in four energy bins give no significant evidence of a change towards heavier or lighter composition with energy.

The results of the composition analysis are naturally limited by the size of the dataset, and by the systematic uncertainties. The elements helium and nitrogen could not be resolved at our level of statistics and systematics; the uncertainty margins show that both statistical and systematic uncertainties play a role here. The accuracy of the fitted proton fraction is mainly statistics-limited, while fitting the iron fraction is mostly systematics-limited for our dataset.

We note that to be able to give e.g. a meaningful lower bound to light particles (p+He), the measured showers at $X_{\max} \gtrsim 800\,\mathrm{g/cm^2}$ are especially important, due to their very high likelihood ratios for protons versus other elements (see Fig. 5.3). Moreover, these likelihood ratios are not as susceptible to systematic errors of about $10\,\mathrm{g/cm^2}$, as would be the case for lower $X_{\max}$. We have measured about 10 of them, but only two survived the selection criteria for a bias-free sample.

A hybrid trigger setup, requiring fewer particle detectors to trigger, in coincidence with a real-time radio signal trigger, would be helpful here. It would have less tendency towards bias from the particle detector requirements, thus decreasing the

need to flag measured showers. An upgrade of the LORA particle detector array, doubling the number of detectors, is planned for the near future. Together with the hybrid trigger, this would allow unbiased $X_{\max}$ measurements also at lower energies, at least down to $10^{16.7}$ eV. This is interesting as it would sample more completely the feature in the cosmic-ray spectrum known as the 'second knee' at, or slightly above, 10^{17} eV. It would also provide more coverage of the energy spectrum in the range of the transition from galactic to extragalactic origin. Also, the number of high-quality showers would increase substantially, improving the composition measurements.

Bibliography

Aab, A. et al. (2014a). Depth of maximum of air-shower profiles at the Pierre Auger Observatory. I. Measurements at energies above $10^{17.8}$ eV. *Phys. Rev. D*, 90:122005.

Aab, A. et al. (2014b). Probing the radio emission from air showers with polarization measurements. *Phys. Rev. D*, 89:052002.

Aab, A. et al. (2016a). Energy estimation of cosmic rays with the Engineering Radio Array of the Pierre Auger Observatory. *Phys. Rev. D*, 93:122005.

Aab, A. et al. (2016b). Nanosecond-level time synchronization of autonomous radio detector stations for extensive air showers. *Journal of Instrumentation*, 11(01):P01018.

Aab, A. et al. (2017). Observation of a large-scale anisotropy in the arrival directions of cosmic rays above 8 10^{18} eV. *Science*, 357(6357):1266–1270.

Aartsen, M. G. et al. (2016). Anisotropy in cosmic-ray arrival directions in the southern hemisphere based on six years of data from the IceCube detector. *The Astrophysical Journal*, 826(2):220.

Abbasi, R. et al. (2013). Icetop: The surface component of icecube. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 700:188 – 220.

Abeysekara, A. U. et al. (2018). Observation of anisotropy of TeV cosmic rays with two years of HAWC. *The Astrophysical Journal*, 865(1):57.

Abraham, J. et al. (2004). Properties and performance of the prototype instrument for the Pierre Auger Observatory. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 523(1):50 – 95.

Abraham, J. et al. (2010). Trigger and aperture of the surface detector array of the Pierre Auger Observatory. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 613(1):29 – 39.

Abreu, P. et al. (2012). Description of atmospheric conditions at the Pierre Auger Observatory using the Global Data Assimilation System (GDAS). *Astroparticle Physics*, 35(9):591 – 607.

Abu-Zayyad, T. et al. (2000). The prototype high-resolution Fly's Eye cosmic ray detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 450(2):253 – 269.

Afanasiev, B. et al. (2003). *Proceedings Tokyo Workshop on Techniques for the study of EHECRs*, page 35.

Agostinelli, S. et al. (2003). Geant4: a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250 – 303.

Ahn, E.-J., Engel, R., Gaisser, T. K., Lipari, P., and Stanev, T. (2009). Cosmic ray interaction event generator SIBYLL 2.1. *Phys. Rev. D*, 80:094003.

Allan, H. (1971). *Progress in elementary particle and cosmic ray physics*, pages 171–302. Number 10. North-Holland Publishing Company.

Allan, H. and Jones, J. (1966). Radio Pulses from Extensive Air Showers. *Nature*, 212:129–131.

Altamimi, Z., Sillard, P., and Boucher, C. (2002). ITRF2000: A new release of the International Terrestrial Reference Frame for earth science applications. *Journal of Geophysical Research (Solid Earth)*, 107:2214.

Alvarez-Muñiz, J., Carvalho, W. R., and Zas, E. (2012). Monte Carlo simulations of radio pulses in atmospheric showers using ZHAireS. *Astroparticle Physics*, 35(6):325 – 341.

Antoni, T. et al. (2005). KASCADE measurements of energy spectra for elemental groups of cosmic rays: Results and open problems. *Astroparticle Physics*, 24(1):1 – 25.

Apel, W. et al. (2010). The KASCADE-Grande experiment. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 620(2):202 – 216.

Apel, W. et al. (2014). The wavefront of the radio signal emitted by cosmic ray air showers. *Journal of Cosmology and Astroparticle Physics*, 2014(09):025.

Ardouin, D. et al. (2009). Geomagnetic origin of the radio emission from cosmic ray induced air showers observed by CODALEMA. *Astroparticle Physics*, 31:192–200.

Askaryan, G. A. (1962). Excess negative charge of the electron-photon shower and coherent radiation originating from It. radio recording of showers under the ground and on the moon. *Journal of the Physical Society of Japan Supplement*, 17:C257.

Athreya, R. (2009). A new approach to mitigation of radio frequency interference in interferometric data. *The Astrophysical Journal*, 696(1):885.

Bauleo, P. and Martino, J. (2009). The dawn of the particle astronomy era in ultra-high-energy cosmic rays. *Nature*, 458:847.

Bell, A. (1978). The acceleration of cosmic rays in shock fronts. *Monthly Notices of the Royal Astronomical Society*, 182:147.

Bellido, J. et al. (2017). Depth of maximum of air-shower profiles at the Pierre Auger Observatory: Measurements above $10^{17.2}$ eV and Composition Implications. *Proceedings of the 35th International Cosmic Ray Conference, Busan, Korea, PoS*, page 506.

Bergmann, T. et al. (2007). One-dimensional hybrid approach to extensive air shower simulation. *Astroparticle Physics*, 26(6):420 – 432.

Bernlöhr, K. (2014). Simulations of detector arrays and the impact of atmospheric parameters. *Proceedings of the Atmospheric Monitoring for High-Energy Astroparticle Detectors (AtmoHEAD) Conference, Saclay (France), June 10-12, 2013.*

Bezyazeekov, P. et al. (2015). Measurement of cosmic-ray air showers with the Tunka Radio Extension (Tunka-Rex). *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equip-*

ment, 802:89 – 96.

Bezyazeekov, P. et al. (2016). Radio measurements of the energy and the depth of the shower maximum of cosmic-ray air showers by Tunka-Rex. *Journal of Cosmology and Astroparticle Physics*, 2016(01):052.

Birch, K. P. and Downs, M. J. (1993). An updated Edlén equation for the refractive index of air. *Metrologia*, 30(3):155.

Blümer, J., Engel, R., and Hörandel, J. R. (2009). Cosmic rays from the knee to the highest energies. *Progress in Particle and Nuclear Physics*, 63(2):293 – 338.

Bohm, G. and Zech, G. (2010). *Introduction to Statistics and Data Analysis for Physicists*. Verlag Deutsches Elektronen Synchrotron.

Buck, A. (1981). New equations for computing vapor pressure and enhancement factor. *Journal of Applied Meteorology*, 20(12):1527–1532.

Buitink, S. et al. (2007). Amplified radio emission from cosmic ray air showers in thunderstorms. *Astronomy & Astrophysics*, 467:385–394.

Buitink, S. et al. (2013). Shower Xmax determination based on LOFAR radio measurements. In *Proceedings of the 33rd International Cosmic Ray Conference, Rio De Janeiro*.

Buitink, S. et al. (2014). Method for high precision reconstruction of air shower $X_{\max}$ using two-dimensional radio intensity profiles. *Phys. Rev. D*, 90:082003.

Buitink, S. et al. (2016). A large light-mass component of cosmic rays at $10^{17} - 10^{17.5}$ electronvolts from radio observations. *Nature*, page 70 .

Cataldi, G. (2017). Towards AugerPrime: the upgrade of the Pierre Auger Observatory. *Nuclear and Particle Physics Proceedings*, 291-293:96 – 101. New eyes on the Universe, CRIS 2016 Cosmic Rays International Seminars Proceedings of the Cosmic Rays International Seminars.

Ciddor, P. E. (1996). Refractive index of air: new equations for the visible and near infrared. *Applied Optics*, 35(9):1566–1573.

Corstanje, A. et al. (2017). The effect of the atmospheric refractive index on the radio signal of extensive air showers. *Astroparticle Physics*, 89:23 – 29.

Corstanje, A., Schellart, P., Nelles, A., et al. (2015). The shape of the radio wavefront of extensive air showers as measured with LOFAR. *Astroparticle Physics*, 61:22–31.

De Domenico, M., Settimo, M., Riggi, S., and Bertin, E. (2013). Reinterpreting the

development of extensive air showers initiated by nuclei and photons. *Journal of Cosmology and Astroparticle Physics*, 2013(07):050.

de Vries, K. D., Scholten, O., and Werner, K. (2013). The air shower maximum probed by Cherenkov effects from radio emission. *Astroparticle Physics*, 45:23 – 27.

de Vries, K. D., van den Berg, A. M., Scholten, O., and Werner, K. (2011). Coherent Cherenkov radiation from cosmic-ray-induced air showers. *Phys. Rev. Lett.*, 107:061101.

Defense Mapping Agency (1987). Department of Defense World Geodetic System 1984. *DMA*, TR 8350.2-B(a).

Dewdney, P. (2015). SKA1 system baseline description v2. *https://www.skatelescope.org/wp-content/uploads/2014/03/SKA-TEL-SKO-0000308_SKA1_System_Baseline_v2_DescriptionRev01-part-1-signed.pdf*.

Engel, R., Heck, D., and Pierog, T. (2011). Extensive air showers and hadronic interactions at high energy. *Annual Review of Nuclear and Particle Science*, 61:467.

Falcke, H. et al. (2005). Detection and imaging of atmospheric radio flashes from cosmic ray air showers. *Nature*, 435:313–316.

Falcke, H. and Gorham, P. (2003). Detecting radio emission from cosmic ray air showers and neutrinos with a digital radio telescope. *Astroparticle Physics*, 19(4):477 – 494.

Fermi, E. (1949). On the origin of the cosmic radiation. *Physical Review*, 75:1169.

Gilliland, T. et al. (1938). Maximum usable frequencies for radio sky-wave transmission. *Journal of Research of the National Bureau of Standards*, 20:627–639.

Grabner, M. and Kvicera, V. (2011). *Atmospheric Refraction and Propagation in Lower Troposphere.*

Greisen, K. (1960). Cosmic Ray Showers. *Annual Review of Nuclear and Particle Science*, 10:63–108.

Greisen, K. (1966). End to the cosmic-ray spectrum? *Physical Review Letters*, 16:748.

Heck, D. et al. (1998). *CORSIKA: a Monte Carlo code to simulate extensive air showers.*

Heitler, W. (1954). *The Quantum Theory of Radiation.* Oxford University Press.

Hess, V. (1912). Durchdringende Strahlung bei sieben Freiballonfahrten. *Physik. Zeitschr. XIII.*

Hillas, A. (1984). The origin of ultra-high energy cosmic rays. *Annual Review of Astronomy and Astrophysics*, 22:425.

Hoover, S. et al. (2010). Observation of ultrahigh-energy cosmic rays with the ANITA balloon-borne radio interferometer. *Phys. Rev. Lett.*, 105:151101.

Hörandel, J. (2003). On the knee in the energy spectrum of cosmic rays. *Astroparticle Physics*, 19(2):193 – 220.

Hörandel, J. (2008). The origin of galactic cosmic rays. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 588(1):181 – 188. Proceedings of the First International Conference on Astroparticle Physics.

Huege, T. (2012). Theory and simulations of air shower radio emission. *AIP Conf. Proc. 5th Int. ARENA Workshop 2012*, 1535:121–127.

Huege, T. (2013). The renaissance of radio detection of cosmic rays. In *Proceedings of the 33rd International Cosmic Ray Conference, Rio De Janeiro.*

Huege, T. (2016). Radio detection of cosmic ray air showers in the digital era. *Physics Reports*, 620:1.

Huege, T. et al. (2012). The LOPES experiment: Recent results, status and perspectives. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 662:S72 – S79. 4th International workshop on Acoustic and Radio EeV Neutrino detection Activities.

Huege, T., Ludwig, M., and James, C. W. (2013). Simulating radio emission from air showers with CoREAS. *ARENA 2012, AIP Conf. Proc. 1535*, pages 128–132.

Huege, T., Ludwig, M., and James, C. W. (2013). Simulating radio emission from air showers with CoREAS. *AIP Conference Proceedings*, 1535(1):128–132.

Ivanov, A., Knurenko, S., and Sleptsov, I. (2009). Measuring extensive air showers with Cherenkov light detectors of the Yakutsk array: the energy spectrum of cosmic rays. *New Journal of Physics*, 11(6):065008.

James, C. W., Falcke, H., Huege, T., and Ludwig, M. (2011). General description of electromagnetic radiation processes based on instantaneous charge acceleration in "endpoints". *Phys. Rev. E*, 84:056602.

Jelley, J. et al. (1965). Radio pulses from extensive cosmic-ray air showers. *Nature*, 205:327.

Kahn, F. D. and Lerche, I. (1966). Radiation from cosmic ray air showers. *Royal Society of London Proceedings Series A*, 289:206–213.

Kamata, K. and Nishimura, J. (1958). The lateral and the angular structure functions of electron showers. *Progress of Theoretical Physics Supplement*, 6:93–155.

Kampert, K.-H. and Unger, M. (2012). Measurements of the cosmic ray composition with air shower experiments. *Astroparticle Physics*, 35(10):660 – 678.

Kazemi, S. et al. (2011). Radio interferometric calibration using the SAGE algorithm. *Monthly Notices of the Royal Astronomical Society*, 414(2):1656–1666.

Knurenko, S. et al. (2015). Mass composition of cosmic rays in the energy region 10^{16} - 10^{18} eV by data the Small Cherenkov Array at Yakutsk. Comparison with other Arrays. *Proceedings of the 34th International Cosmic Ray Conference, The Hague, The Netherlands, PoS*, page 254.

Marin, V. and Revenu, B. (2012). Simulation of radio emission from cosmic ray air shower with SELFAS2. *Astroparticle Physics*, 35(11):733 – 741.

Matthews, J. (2005). A Heitler model of extensive air showers. *Astroparticle Physics*, 22:387.

Matthews, J. N. et al. (2011). The Telescope Array Experiment. In *Proceedings, 32nd International Cosmic Ray Conference (ICRC 2011): Beijing, China, August 11-18, 2011*, volume 2, page 273.

Mitra, P. et al. (2019). Correcting atmospheric effects on radio emission from cosmic rays with GDAS. *in preparation*.

Mulrey, K. et al. (2019). Calibration of the LOFAR low-band antennas using the Galaxy and a model of the signal chain. *accepted by Astroparticle Physics (2019)*. Arxiv:1903.05988.

National Oceanic and Atmospheric Administration (2010). Global Data Assimilation System (GDAS). *www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-data-assimilation-system-gdas*.

Nelles, A. et al. (2015a). Calibrating the absolute amplitude scale for air showers measured at LOFAR. *Journal of Instrumentation*, 10(11):P11005.

Nelles, A. et al. (2015b). A parameterization for the radio emission of air showers as predicted by CoREAS simulations and applied to LOFAR measurements.

Astroparticle Physics, 60:13 – 24.

Nelles, A. et al. (2015c). The radio emission pattern of air showers as measured with LOFAR: a tool for the reconstruction of the energy and the shower maximum. *Journal of Cosmology and Astroparticle Physics*, 2015(05):018.

Nelles, A., Schellart, P., et al. (2015d). Measuring a Cherenkov ring in the radio emission from air showers at 110 - 190 MHz with LOFAR. *Astroparticle Physics*, 65:11 – 21.

Nigl, A. et al. (2008). Direction identification in radio images of cosmic-ray air showers detected with LOPES and KASCADE. *Astronomy & Astrophysics*, 487:781–788.

Offringa, A. R. et al. (2010). Post-correlation radio frequency interference classification methods. *Monthly Notices of the Royal Astronomical Society*, 405(1):155–167.

Offringa, A. R. et al. (2013). The LOFAR radio environment. *Astronomy & Astrophysics*, 549:A11.

Ostapchenko, S. (2013). QGSJET-II: physics, recent improvements, and results for air showers. *EPJ Web of Conferences*, 52:02001.

Papoulis, A. and Pillai, S. (2002). *Probability, Random Variables and Stochastic Processes, 4th ed.* McGraw-Hill.

Pearson, T. J. and Readhead, A. C. S. (1984). Image formation by self-calibration in radio astronomy. *Ann. Rev. Astron. Astrophys.*, 22:97–130.

Pierog, T., Karpenko, I., Katzy, J. M., Yatsenko, E., and Werner, K. (2015). EPOS LHC: test of collective hadronization with data measured at the CERN Large Hadron Collider. *Phys. Rev. C*, 92:034906.

Prosin, V. et al. (2014). Tunka-133: Results of 3 year operation. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 756:94 – 101.

Prosin, V. et al. (2015). Primary CR energy spectrum and mass composition by the data of Tunka-133 array. *EPJ Web of Conferences*, 99:04002.

Ravel, O. (2012). The CODALEMA experiment. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 662:S89 – S94. 4th International workshop on Acoustic and Radio EeV Neutrino detection Activities.

Rayleigh (1905). The problem of the random walk. *Nature*, 72:318.

Royal Netherlands Meteorological Institute (KNMI) (2011). Dutch Climate Atlas.

Rüeger, J. (2002). Refractive index formulae for radio waves. *Proceedings of FIG XXII International Congress, 2002*.

Schellart, P. et al. (2014). Polarized radio emission from extensive air showers measured with LOFAR. *Journal of Cosmology and Astroparticle Physics*, 2014(10):014.

Schellart, P., Nelles, A., et al. (2013). Detecting cosmic rays with the LOFAR radio telescope. *Astronomy & Astrophysics*, 560:A98.

Schellart, P., Trinh, T. N. G., et al. (2015). Probing atmospheric electric fields in thunderstorms through radio emission from cosmic-ray-induced air showers. *Phys. Rev. Lett.*, 114:165001.

Schmidt, A. et al. (2011). FPGA based signal-processing for radio detection of cosmic rays. *Nuclear Science, IEEE Transactions on*, 58(4):1621–1627.

Scholten, O., de Vries, K. D., and Werner, K. (2012). Coherent radiation from extensive air showers. *Nuclear Instruments and Methods in Physics Research Section A*, 662, Supplement 1:S80 – S84. 4th International workshop on Acoustic and Radio EeV Neutrino detection Activities.

Scholten, O., Trinh, T. N. G., et al. (2016). Measurement of the circular polarization in radio emission from extensive air showers confirms emission mechanisms. *Phys. Rev. D*, 94:103010.

Scholten, O., Werner, K., and Rusydi, F. (2008). A macroscopic description of coherent geo-magnetic radiation from cosmic-ray air showers. *Astroparticle Physics*, 29:94–103.

Schröder, F. et al. (2010). New method for the time calibration of an interferometric radio antenna array. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 615(3):277 – 284.

Schröder, F. G. (2017). Radio detection of cosmic-ray air showers and high-energy neutrinos. *Progress in Particle and Nuclear Physics*, 93:1 – 68.

Schröder, F. G. et al. (2011). Investigation of the Radio Wavefront of Air Showers with LOPES and REAS3. In *International Cosmic Ray Conference*, volume 3 of *International Cosmic Ray Conference*, page 64.

Schröder, F. G. et al. (2012a). Cosmic ray measurements with LOPES: Status and recent results. *AIP Conf. Proc. 5th Int. ARENA Workshop 2012*, 1535:78.

Schröder, F. G. et al. (2012b). On noise treatment in radio measurements of cosmic ray air showers. *Nuclear Instruments and Methods in Physics Research A*, 662:238.

Schulz, J. et al. (2015). Status and prospects of the Auger Engineering Radio Array. In *Proceedings, 34nd International Cosmic Ray Conference (ICRC 2015): The Hague, The Netherlands, July 30 - August 6, 2015*, volume 236, page 615.

Smida, R. et al. (2014). First experimental characterization of microwave emission from cosmic ray air showers. *Phys. Rev. Lett.*, 113:221101.

Sokolsky, P. (2011). Final results from the High solution Fly's Eye (HiRes) Experiment. *Nuclear Physics B - Proceedings Supplements*, 212-213:74 – 78. Proceedings of the Cosmic Ray International Seminars (CRIS 2010).

Stanev, T. (2010). *High energy cosmic rays*. Springer.

Szadkowski, Z., Fraenkel, E., and van den Berg, A. (2013). FPGA/NIOS implementation of an adaptive FIR filter using linear prediction to reduce narrow-band RFI for radio detection of cosmic rays. *Nuclear Science, IEEE Transactions on*, 60(5):3483–3490.

Taylor, G. B., Carilli, C. L., and Perley, R. A. (1999). *Synthesis Imaging in Radio Astronomy II*, volume 180.

Thoudam, S. et al. (2014). LORA: A scintillator array for LOFAR to measure extensive air showers. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 767:339 – 346.

Thoudam, S. et al. (2016). Measurement of the cosmic-ray energy spectrum above 10^{16} eV with the LOFAR Radboud Air Shower Array. *Astroparticle Physics*, 73:34 – 43.

Trinh, T. N. G., Scholten, O., et al. (2016). Influence of atmospheric electric fields on the radio emission from extensive air showers. *Phys. Rev. D*, 93:023003.

Trinh, T. N. G., Scholten, O., et al. (2017). Thunderstorm electric fields probed by extensive air showers through their polarized radio emission. *Phys. Rev. D*, 95:083004.

U.S. Government Printing Office (1976). U.S. Standard Atmosphere, 1976.

van Haarlem, M. P. et al. (2013). LOFAR: The Low Frequency Array. *Astronomy and Astrophysics*, 556:A2.

Werner, K., de Vries, K. D., and Scholten, O. (2012). A realistic treatment of geomagnetic Cherenkov radiation from cosmic ray air showers. *Astroparticle Physics*, 37:5 – 16.

Wessels, H. (1998). Evaluation of a radio interferometry lightning positioning system. Technical report, KNMI Scientific Report.

Wijnholds, S., van der Tol, S., Nijboer, R., and van der Veen, A.-J. (2010). Calibration challenges for future radio telescopes. *Signal Processing Magazine, IEEE*, 27(1):30–42.

Zatsepin, G. and Kuzmin, V. (1966). Upper limit of the spectrum of cosmic rays. *Journal of Experimental and Theoretical Physics Letters*, 4:78.

Zimmermann, B. et al. (2017). Cosmic ray physics with the Auger Engineering Radio Array (AERA). *Nuclear and Particle Physics Proceedings*, 291-293:90 – 95. New eyes on the Universe, CRIS 2016 Cosmic Rays International Seminars Proceedings of the Cosmic Rays International Seminars.

Summary

Cosmic rays continuously arrive on Earth from all directions. These high-energy particles, with energies spanning a range from 10^9 to over 10^{20} eV, are known to be atomic nuclei. Although much has been learned since their discovery in 1912, significant questions are still open, regarding their origin and acceleration, their propagation through space, and their mass composition. The answers to these questions tell us about the objects in the universe capable of accelerating particles up to the highest energies. We know these objects must either have rather extreme conditions, or be very large, to produce particles many orders of magnitude more energetic than those in man-made accelerators. Cosmic rays are therefore messengers from these places in the universe, and measuring those particles is in this respect similar to astronomy, where we measure signals spanning the entire electromagnetic spectrum to image and learn about far-away objects. The field of *astroparticle physics* is an extension of astronomy to include other information sources, such as cosmic rays, neutrinos, the highest-energy gamma rays, and perhaps dark matter, if detections can be made.

Above about 10^{14} eV, the flux of cosmic rays is too small for direct detection in balloons or satellites, and ground-based observatories then provide the large surface area that is needed. A cosmic ray particle interacts with nuclei inside air molecules

in the atmosphere, to produce secondary particles. As these also interact, a cascade of secondary particles called *extensive air shower* travels towards the ground and can be detected there. The number of particles reaches a maximum at a depth in the atmosphere referred to as $X_{\max}$. This indirect detection can no longer identify the mass of an individual primary particle. However, for a (preferably large) sample of detections, the mass composition can be estimated statistically, from the 'footprints' they leave in the detectors. This is possible because on average, $X_{\max}$ varies with the type of primary particle.

The air showers produce radio signals, arriving on the ground as short pulses, best detected in the range of 30 to 80 MHz, close to the frequencies of public FM transmitters. Radio antennas and digital receivers are relatively cheap, and can work essentially all the time, except during thunderstorms which change the radio signals through their electric fields.

The radio detection technique has evolved over the last 15 years, from an experimental technique to a proven, complementary measurement mode along particle detectors and telescopes. Part of this development has been done using the LOFAR (Low Frequency ARray) radio telescope in the Netherlands. It features a core region with a high density of antennas operating in the 30 to 80 MHz frequency range, making it very useful to obtain densely sampled measurements of the radio signals. LOFAR has detected its first cosmic ray on June 12, 2011, and up to late 2018 has produced 2717 successful detections, of which 702 were strong enough to be suitable for further, high-precision analysis.

This thesis also concerns radio measurements of cosmic rays with LOFAR. The aim of this work is to show the high accuracy to which the radio signals can be measured, and to further improve the accuracy of the cosmic-ray measurements, including establishing their mass composition. As the radio pulse is short (10 to 100 nanoseconds), it arrives at the various antennas on the ground in a distinct, connected wavefront. As the signal originates from a finite distance, the wavefront is not flat. In Chapter 2, measurements of the wavefront are presented, by timing individual pulses to a precision better than 1 nanosecond. Of the various proposed wavefront shapes, such as a sphere, cone, or hyperboloid, the hyperboloid shape was found to fit best. No structural deviations from the fit were seen, hence the hyperboloid parametrization is sufficient to this level of precision.

Measuring the wavefront relies on accurate timing calibration of the antennas and their data acquisition. To this end, a calibration method was developed, relying on public FM transmitters which were always on and at a known location (presented in Chapter 3). Measuring the relative phases of those radio waves produces a timing calibration, as long as deviations are smaller than one wavelength. Realizing that the phase differences across antennas of these transmitters are stable, while those from galactic 'noise' are not, it turns out to be useful to measure the phase stability versus frequency. This delivered a practical, and previously unexplored method to clean the frequency spectrum of our data from man-made transmitters.

To infer the position of the shower maximum ($X_{\max}$), the measured intensity footprints on the ground are compared to those from detailed simulations of the radio signal from air showers. This has been demonstrated earlier to give a precision of $20\,\mathrm{g/cm^2}$, on typical values of $X_{\max}$ between 550 and $850\,\mathrm{g/cm^2}$, which is comparable to e.g. fluorescence detection. However, for high-precision work, it is important to address systematic uncertainties. One of the largest was given by the variations in the atmosphere above LOFAR. In Chapter 4 it is shown that the variations in density profile and in the refractive index need to be fully taken into account, to remove a systematic uncertainty of 4 to $11\,\mathrm{g/cm^2}$ in $X_{\max}$. The refractive index determines the Cherenkov angle, and a toy model shows that the variations in the Cherenkov angle roughly describe the systematic error on $X_{\max}$, when not taken into account. After the simulation software was adapted, the local atmosphere is now fully accounted for.

Finally, Chapter 5 presents the analysis of the mass composition of our data sample, acquired between 2011 and 2018. Systematic uncertainties are now smaller, from the improved treatment of the atmosphere and from inferring the primary energy from radio data instead of the (relatively small) particle detector array. Performing the footprint fit on radio data only, instead of combining radio and particle data, is another step towards a stand-alone radio measurement; the particle detectors are used merely as a trigger for the radio readout. The data selection criteria have been improved, in order to obtain a bias-free sample of the $X_{\max}$-distribution. Using a maximum-likelihood based statistical analysis, the composition was found to be mostly intermediate-mass particles, ranging from helium to oxygen (or slightly beyond), at a best fit of 47 to 78 %, depending on the choice of hadronic interaction

model. Only a minority of the cosmic particles are protons, in our energy range. Their fraction is just below 20 % at best fit, and has an upper bound of 43 % (99 % confidence level) which holds for all three hadronic interaction models considered. It has been shown that with radio measurements, the air shower parameters including X_{max} can be accurately inferred. A mass composition analysis has been demonstrated, from analyzing only radio data, triggered by particle detectors. Systematic and statistical uncertainties on X_{max} are now in line with existing state-of-the-art methods. Therefore, the radio method is well suited either as a complementary technique for existing arrays, or as the main measurement technique alongside a smaller particle detector array as is the case at LOFAR.

Samenvatting

Kosmische straling bestaat uit geladen deeltjes die voortdurend vanuit alle richtingen op aarde aankomen. Deze deeltjes, meest atoomkernen, kunnen een enorm bereik aan energie hebben, van relatief lage tot extreem hoge energie, ruim een miljoen maal wat in aardse deeltjesversnellers bereikt wordt. Die laatste zijn dan ook de hoogst energetische deeltjes die we kennen.

Ondanks veel onderzoek zijn er nog belangrijke vragen open, zowel over hun oorsprong, hun reis door de ruimte, als over het type deeltjes. Antwoorden op deze vragen vertellen ons iets over de objecten in het heelal die deze deeltjes tot de hoogste energie kunnen versnellen. We weten dat in deze objecten extreme omstandigheden moeten heersen, of dat ze heel groot moeten zijn. Dit maakt de kosmische deeltjes tot boodschappers van deze plaatsen in het heelal. Het meten van de deeltjes vormt hiermee een uitbreiding van de (traditionele) astronomie, waarin men licht en andere elektromagnetische signalen gebruikt om afbeeldingen te maken en te leren over ver weg gelegen objecten. In het algemeen wordt dit vakgebied aangeduid als *astrodeeltjesfysica*, en omvat ook neutrino's, gammastraling van zeer hoge energie, en wellicht donkere materie, als deze gedetecteerd kan worden.

Boven een energie van ca. 10^{14} eV zijn de kosmische deeltjes te zeldzaam voor directe detectie met behulp van ballonnen of satellieten. Ze zijn echter wel indirect

te meten op de grond. Een deeltje dat de atmosfeer binnenkomt, zal botsen op een atoomkern in een luchtmolecuul. Hierbij ontstaan nieuwe deeltjes, die eveneens zullen botsen. Dit produceert een *deeltjeslawine* (Engels: *extensive air shower*) die op de grond gedetecteerd kan worden. Het aantal deeltjes bereikt een maximum, op een diepte in de atmosfeer aangeduid als $X_{\max}$, en daalt daarna weer.

De deeltjeslawines produceren radiosignalen, die als een korte puls op de grond aankomen. Ze zijn het best te meten op frequenties beneden 100 MHz, wat rond de frequentie van FM-radiozenders ligt. De radio-meettechniek van kosmische straling heeft zich ontwikkeld over de afgelopen ca. 15 jaar, van een experimentele techniek tot een bewezen complementaire meetmethode naast deeltjesdetectors en telescopen. Een deel van deze ontwikkeling is gebeurd met de LOFAR (LOw Frequency ARray) radio telescoop in Drenthe. In het centrale gebied heeft deze enkele honderden antennes die meten tussen 30 en 80 MHz, waarmee hoge-resolutie opnamen gemaakt kunnen worden. LOFAR heeft zijn eerste kosmische deeltje gedetecteerd op 12 juni 2011, en tot eind 2018 in totaal 2717, waarvan 702 een goed genoeg signaal hadden voor een verdere, hoge-precisie analyse.

Dit proefschrift gaat over radio-metingen van kosmische straling met LOFAR. Doel van dit werk is om te laten zien hoe nauwkeurig het radiosignaal gemeten kan worden en, in het verlengde, om de nauwkeurigheid van de metingen aan kosmische deeltjes te verbeteren, inclusief het bepalen van hun massa-compositie.

Aangezien de radiopulsen kort zijn (10 tot 100 nanoseconden), komen ze aan bij de antennes als een gesloten golffront met een dikte van enkele meters. Het signaal komt van een eindige afstand in de atmosfeer (enkele kilometers), waardoor het geen vlakke golf is. Hoofdstuk 2 gaat over metingen van het golffront, door de pulsen te timen tot op 1 nanoseconde nauwkeurig. Van de voorgestelde vormen van het golffront, zoals een deel van een bol, een kegel of een hyperboloide, blijkt de hyperboloide het best te passen. Ook waren (tot op de meetnauwkeurigheid) geen afwijkingen van deze vorm zichtbaar.

Om het golffront te kunnen meten moeten de signalen van de antennes precies gesynchroniseerd worden. Om dit voor elke meting te kunnen doen, is een calibratie-methode ontwikkeld (hoofdstuk 3), gebaseerd op signalen van publieke FM-zenders, aangezien die altijd aan staan en vanaf een bekende plaats uitzenden. Het faseverschil tussen twee antennes moet dan een bekende, vaste waarde hebben waarmee

klokverschillen gecorrigeerd kunnen worden, zolang deze kleiner zijn dan één golflengte. Op frequenties waar geen zender uitzendt, is slechts 'ruis' uit de Melkweg, waardoor de faseverschillen willekeurig zijn. Het blijkt nuttig om de fase-stabiliteit voor alle frequenties te bepalen, en zo alle menselijke (stoor)zenders uit onze data te filteren; een praktische methode die nog niet eerder onderzocht was.

Om de positie van het maximum van de deeltjeslawine $X_{\max}$ te bepalen, wat cruciaal is voor het bepalen van de massacompositie, vergelijken we de gemeten signaalsterkte op de grond met gedetailleerde simulaties van het radiosignaal. Zoals al eerder was gebleken, kan hiermee een precisie van $20\,\mathrm{g/cm^2}$ bereikt worden, op typische waarden van $X_{\max}$ van 550 tot ca. $850\,\mathrm{g/cm^2}$. Dit is vergelijkbaar is met de best beschikbare andere methoden, bv. fluorescentie-detectie. Echter, voor hogeprecisie metingen is het ook belangrijk om systematische afwijkingen te beperken. Een van de grootste hiervan werd gegeven door de variaties in de atmosfeer boven LOFAR. Het dichtheidsprofiel en de brekingsindex van de atmosfeer zijn hierbij het belangrijkste, zoals in hoofdstuk 4 wordt uiteengezet. De simulatiesoftware is zo aangepast dat de atmosferische gegevens op het moment van de meting meegenomen worden, wat een afwijking in de gemiddelde gemeten $X_{\max}$ van 4 tot $11\,\mathrm{g/cm^2}$ voorkomt.

In hoofdstuk 5 is de massa-compositie van kosmische straling bepaald in ons energiegebied, iets boven $10^{17}\,\mathrm{eV}$. Naast het meenemen van de atmosferische variaties zijn nog diverse verbeteringen doorgevoerd. Zo wordt de energie van het deeltje nu afgeleid uit de sterkte van het radiosignaal in plaats van uit de (veel kleinere) set deeltjesdetectors. Het vergelijken van simulaties en meetwaarden wordt nu ook alleen op basis van radiosignalen gedaan, wat een verdere stap naar zelfstandige radiometingen vormt; de deeltjesdetectors worden nu alleen als trigger gebruikt om het radiosignaal uit te lezen. De selectiecriteria voor de data zijn verbeterd, om een waarheidsgetrouwe meting van de $X_{\max}$-verdeling te verkrijgen.

Statistische analyse laat zien dat de kosmische deeltjes bij deze energie voor meer dan de helft bestaan uit atoomkernen met middelmatige massa, vanaf helium ongeveer tot en met zuurstof. De lichtste deeltjes, protonen, zijn in de minderheid; hun aandeel is maximaal 43 %, met hoge statistische significantie, en onafhankelijk van de 3 gebruikte modellen van deeltjesinteracties.

Door radiometingen kunnen de eigenschappen van deeltjeslawines inclusief de

positie van hun maximum goed bepaald worden. Een bepaling van de massaverdeling kon gedemonstreerd worden, op basis van alleen radiometingen, getriggerd door deeltjesdetectors. Statistische en systematische onzekerheden in $X_{\max}$ komen nu overeen met die van bestaande state-of-the-art technieken. Dit maakt de radiomethode zeer geschikt, ofwel als complementaire techniek naast bestaande meetexperimenten, ofwel als de hoofdtechniek naast een kleinere set deeltjesdetectors, zoals bij LOFAR.

List of publications

Peer-reviewed

A. Corstanje, P. Schellart, A. Nelles et al. (LOFAR Collaboration)
The shape of the radio wavefront of extensive air showers as measured with LOFAR
Astroparticle Physics 61, 22 (2015)

A. Corstanje et al. (LOFAR Key Science Project Cosmic Rays)
Timing calibration and spectral cleaning of LOFAR time series data
Astronomy & Astrophysics 590, A41 (2016)

A. Corstanje et al. (LOFAR Key Science Project Cosmic Rays)
The effect of the atmospheric refractive index on the radio signal of extensive air showers
Astroparticle Physics 89, 23 (2017)

P. Schellart, A. Nelles et al. (LOFAR Collaboration)
Detecting cosmic rays with the LOFAR radio telescope
Astronomy & Astrophysics 560, A98 (2013)

P. Schellart et al. (LOFAR Key Science Project Cosmic Rays)
Polarized radio emission from extensive air showers measured with LOFAR
Journal of Cosmology and Astroparticle Physics 10, 14 (2014)

P. Schellart et al. (LOFAR Key Science Project Cosmic Rays)
Polarized radio emission from extensive air showers measured with LOFAR
Journal of Cosmology and Astroparticle Physics 10, 14 (2014)

S. Thoudam et al. (LOFAR Key Science Project Cosmic Rays)
LORA: A scintillator array for LOFAR to measure extensive air showers
Nuclear Instruments and Methods in Physics Research Section A 767, 339 (2014)

S. Buitink et al. (LOFAR Key Science Project Cosmic Rays)
Method for high precision reconstruction of air shower $X_{\max}$ using two-dimensional radio intensity profiles
Physical Review D 90, 082003 (2014)

A. Nelles, P. Schellart et al. (LOFAR Collaboration)
Measuring a Cherenkov ring in the radio emission from air showers at 110–190 MHz with LOFAR
Astroparticle Physics 65, 11 (2015)

A. Nelles et al. (LOFAR Key Science Project Cosmic Rays)
The radio emission pattern of air showers as measured with LOFAR: a tool for the reconstruction of the energy and the shower maximum
Journal of Cosmology and Astroparticle Physics 05, 018 (2015)

P. Schellart, T.N.G. Trinh et al. (LOFAR Collaboration)
Probing Atmospheric Electric Fields in Thunderstorms through Radio Emission from Cosmic-Ray-Induced Air Showers
Physical Review Letters 114, 165001 (2015)

S. Thoudam et al. (LOFAR Key Science Project Cosmic Rays)
Measurement of the cosmic-ray energy spectrum above 10^{16} eV *with the LOFAR Radboud Air Shower Array*
Astroparticle Physics 73, 34 (2016)

A. Nelles, J.R. Hörandel, T. Karskens, M. Krause et al. (LOFAR Collaboration)
Calibrating the absolute amplitude scale for air showers measured at LOFAR
Journal of Instrumentation 10, P11005 (2015)

T.N.G. Trinh, O. Scholten, S. Buitink et al. (LOFAR Key Science Project Cosmic Rays)
Influence of atmospheric electric fields on the radio emission from extensive air showers
Physical Review D 93, 023003 (2016)

S. Buitink et al. (LOFAR Collaboration)
A large light-mass component of cosmic rays at 10^{17} - $10^{17.5}$ eV *from radio observations*
Nature 531, 70 (2016)

O. Scholten, T.N.G. Trinh et al. (LOFAR Key Science Project Cosmic Rays)
Measurement of the circular polarization in radio emission from extensive air showers confirms emission mechanisms
Physical Review D 94, 103010 (2016)

T.N.G.Trinh, O. Scholten et al. (LOFAR Key Science Project Cosmic Rays)
Thunderstorm electric fields probed by extensive air showers through their polarized radio emission
Physical Review D 95, 083004 (2017)

B.M. Hare et al. (LOFAR Key Science Project Cosmic Rays)
LOFAR lightning imaging: Mapping lightning with nanosecond precision
Journal of Geophysical Research: Atmospheres 123, 2861 (2018)

K. Mulrey et al. (LOFAR Key Science Project Cosmic Rays)
Calibration of the LOFAR low-band antennas using the Galaxy and a model of the signal chain
Astroparticle Physics 111, 1 (2019)

B. Hare et al. (LOFAR Key Science Project Cosmic Rays)
Needle-like structures discovered on positively charged lightning branches
Nature 568, 360 (2019)

P. Mitra et al. (LOFAR Key Science Project Cosmic Rays)
Correcting atmospheric effects on radio emission from cosmic rays with GDAS
submitted to Astroparticle Physics (2019)

Conference proceedings, first author

A. Corstanje et al. (LOFAR Collaboration)
LOFAR: Detecting cosmic rays with a radio telescope
Proceedings of the 32nd International Cosmic Ray Conference 2011, Beijing, China

A. Corstanje, S. Buitink et al. (LOFAR Key science project Cosmic Rays)
The influence of the atmospheric refractive index on radio Xmax measurements of air showers
7th International Conference on Acoustic and Radio EeV Neutrino Detection Activities (ARENA 2016)
EPJ Web of Conferences 135, 01012 (2017)

A. Corstanje et al. (LOFAR Key science project Cosmic Rays)
Polarized radio emission and radio wavefront shape of extensive air showers

Proceedings of the 34th International Cosmic Ray Conference 2015, The Hague, The Netherlands

Publications as member of the LOFAR Collaboration

M. van Haarlem et al. (LOFAR Collaboration)
LOFAR: The LOw-Frequency ARray
Astronomy & Astrophysics 556, A2 (2013)

J.B.R. Oonk et al. (LOFAR Collaboration)
Discovery of carbon radio recombination lines in absorption towards Cygnus A
Monthly Notices of the Royal Astronomical Society 437, 3506 (2014)

H.K. Vedantham et al. (LOFAR Collaboration)
Lunar occultation of the diffuse radio sky: LOFAR measurements between 35 and 80 MHz
Monthly Notices of the Royal Astronomical Society 450, 2291 (2015)

T. Coenen et al. (LOFAR Collaboration)
The LOFAR pilot surveys for pulsars and fast radio transients
Astronomy & Astrophysics 570, A60 (2014)

C. Sobey et al. (LOFAR Collaboration)
LOFAR discovery of a quiet emission mode in PSR B0823+26
Monthly Notices of the Royal Astronomical Society 451, 2493 (2015)

J.N. Girard et al. (LOFAR Collaboration)
Imaging Jupiter's radiation belts down to 127 MHz with LOFAR
Astronomy & Astrophysics 587, A3 (2016)

A.J. Stewart et al. (LOFAR Collaboration)

LOFAR MSSS: detection of a low-frequency radio transient in 400 h of monitoring of the North Celestial Pole
Monthly Notices of the Royal Astronomical Society 456, 2321 (2016)

N. Jackson et al. (LOFAR Collaboration)
LBCS: The LOFAR Long-Baseline Calibrator Survey
Astronomy & Astrophysics 595, A86 (2016)

Acknowledgements

This thesis is the end product of a long journey. A decade ago, LOFAR was only in its testing phase, and getting any data (let alone cosmic ray triggers) from its transient buffers was still quite a challenge. Our LOFAR Cosmic Rays group has assisted Astron in getting our observing mode to work, installed the LORA particle detectors, and recorded its first cosmic ray signal on June 12, 2011. A data analysis pipeline was written, starting from literally a blank screen. Many analysis steps followed in the years after, leading to several published papers such as those listed in Sect. 1.7, and to which this thesis is a contribution.

It has been a joy, and I am grateful to have been part of this group throughout the various phases. Along the way I have met many people I would like to thank. A list of everyone involved would be rather long, but at least the following people should be mentioned in particular.

First, my promotor Heino Falcke, for starting me on this project, for his vision and guidance throughout this work. Also, for his confidence and patience when I got stuck. My second promotor Jörg Hörandel, who had a central role in managing and supervising our group, including my own work. The (group) discussions on all those smaller and larger issues one encounters in research were very important for progress.

The Nijmegen astronomy department as a whole (including all those not explicitly named here), which hosts a variety of people of at least 20 nationalities. It combines a positive, friendly atmosphere with getting serious work done, and I have felt at home here doing this research. This is very valuable, and something I do not just take for granted. The various social activities and traditions, for instance the cakes for birthdays and published papers, the Christmas dinners, playing table football, journal clubs etc. all help to make people feel at home and have a lively experience here.

Andreas and Clancy, thanks for getting me started, Sander and Satyendra for the discussions in getting to know the inner workings of LOFAR and LORA, not to mention for enabling us to get data in the first place. It was nice to be part of this, see how things work on the inside, and to help out in getting the system going.

Of course, the vital role of Pim and Anna needs to be mentioned, how they laid the foundations of success by producing pipeline software that still works today (as long as we are careful not to mess it up...) and is still regarded as the standard of our analysis. Most of all, their light-hearted attitude while solving tough problems, has been a real inspiration for me. When walking by their office, just a short discussion (and some of that remarkable sense of humour) would often recharge me to tackle my own research issue of the day.

Stijn, for his sharp eye on cosmic ray analysis in general, and on the composition measurement in particular. This has helped especially when doing the final chapter. Also, the others forming 'the muscles from Brussels': Katie, Tobi, Pragati, Hershal, Godwin, who have helped with timely and sometimes difficult improvements to our measurements. Laura and Antonio, for some of that Italian spirit, and keeping the group lively when people dispersed.

Olaf, Gia and Brian from Groningen, and Jörg Rachen from Nijmegen, whose discussions from a somewhat unusual angle are often helpful in getting the details right.

Tim Huege and Tanguy Pierog from Karlsruhe, for discussions on the inner workings of the CoREAS simulation software and for producing the improved version used in Chapter 5.

Our (former) master students Maria, Tijs, and Ilse, for their hard and enthousiastic work on the surprisingly complex task of calibrating the LOFAR antennas,

and on a functional description of the radio footprints.

Daan, for the computer support, and the reserved time for a simulation job that was rather big for our computer cluster. Martin, for the LaTeX template for this thesis.

Thanks to the people at Astron for their support in getting our LOFAR observing mode to work, and for putting up with all our questions, typically requesting something different from the planned astronomical data flow... Particular names that come to mind are Pieter Donker, Ruud Overeem, André Gunst, Menno Norden, Arie Doorduin, Stefan Wijnholds, Michael Wise, Roberto Pizzo, Wilfred Frieswijk, Rebecca McFadden, Maaijke Mevius, and in recent years also (again) Sander ter Veen.

Sarka, thanks for our ongoing discussions about life in science, and life in general, while more than once reminding me of the true scientific spirit.

Finally, the importance of family support is not to be underestimated; in particular, thanks to my parents and brother for their continuous support during this work.